

PREDICCIÓN A CORTO PLAZO DE LA DEMANDA DE ENERGÍA ELÉCTRICA USANDO PROCESOS GAUSSIANOS (GP)

JORGE ELIÉCER LOAIZA MUÑOZ



UNIVERSIDAD TECNOLÓGICA DE PEREIRA

FACULTAD DE INGENIERÍAS

PROGRAMA DE MAESTRÍA EN INGENIERÍA DE SISTEMAS Y
COMPUTACIÓN

PEREIRA

2019

PREDICCIÓN A CORTO PLAZO DE LA DEMANDA DE ENERGÍA
ELÉCTRICA USANDO PROCESOS GAUSSIANOS (GP)

JORGE ELIÉCER LOAIZA MUÑOZ

Proyecto de grado presentado como requisito para optar por el título de
Magíster en Ingeniería de Sistemas y Computación

Director: Carlos David Zuluaga, PhD

UNIVERSIDAD TECNOLÓGICA DE PEREIRA
FACULTAD DE INGENIERÍAS
PROGRAMA DE MAESTRÍA EN INGENIERÍA DE SISTEMAS Y
COMPUTACIÓN
PEREIRA
2019

Nota de aceptación

Firma del presidente del jurado

AGRADECIMIENTOS

A todos aquellos docentes de las diferentes asignaturas de la Maestría, los cuales transmitieron toda la información necesaria para obtener el conocimiento que hizo posible el desarrollo de mi propuesta. Al PhD Carlos David Zuluaga, mi asesor, quien estuvo al tanto de mis dudas e inquietudes.

TABLA DE CONTENIDO

	Pág.
LISTA DE FIGURAS	VIII
LISTA DE TABLAS	IX
RESUMEN	1
ABSTRACT	2
1. INTRODUCCIÓN	3
1.1. JUSTIFICACIÓN	3
1.2. PLANTEAMIENTO DEL PROBLEMA	5
1.3. ANTECEDENTES	7
1.4. OBJETIVOS	10
2. PROCESOS GAUSSIANOS	14
2.1. Función media	15
2.2. Función de covarianza	15
2.2.1. Exponencial Cuadrática	16
2.2.2. Periódica	17
2.3. El Conjunto de datos de entrenamiento	17
2.4. Análisis de regresión	18
2.5. Procesos gaussianos para regresión	19
2.6. Matriz de covarianza	20
2.6.1. Covarianza del conjunto de datos de entrenamiento	20

2.6.2.	Covarianza del conjunto de datos de prueba	21
2.6.3.	Covarianza del conjunto de entrenamiento y conjunto de prueba	21
2.7.	Hiperparámetros	23
2.7.1.	Determinación de los hiperparámetros	23
2.8.	Composición de funciones de covarianza	25
3.	MATERIALES Y MÉTODOS	28
3.1.	Origen de los datos	28
3.2.	Depurado de los datos	29
3.2.1.	Diario	29
3.2.2.	Horario	30
3.3.	Métricas de rendimiento	31
3.3.1.	Error medio cuadrático estandarizado	31
3.3.2.	Media estandarizada del logaritmo de las pérdidas	32
3.4.	Función de covarianza propuesta	32
3.5.	Otros métodos de predicción	33
4.	RESULTADOS EXPERIMENTALES	36
4.1.	Resultados predicción diaria	36
4.2.	Resultados predicción horaria	40
4.3.	Métodos alternativos a los GP para predicción	43
4.4.	Predicción de demanda usando diferentes conjuntos de datos	45
5.	CONCLUSIONES	48
	REFERENCIAS	50

LISTA DE FIGURAS

	Pág.
1.1. <i>Consumo de energía eléctrica en Colombia entre los años 2003 y 2016.</i>	13
2.1. <i>Ejemplo datos de entrenamiento. La barra representa el ruido de cada observación</i>	19
4.1. <i>Estimación de la semana 44 del año 2015. La línea roja representa el consumo real. La línea azul representa las estimaciones.</i>	37
4.2. <i>Estimación diaria con la función de covarianza base SE. La línea roja representa el consumo real. La línea azul representa las estimaciones.</i>	38
4.3. <i>Estimación diaria con la función de covarianza base PER. La línea roja representa el consumo real. La línea azul representa las estimaciones.</i>	39
4.4. <i>Estimación diaria con la función de covarianza base RQ. La línea roja representa el consumo real. La línea azul representa las estimaciones.</i>	40
4.5. <i>Predicción para el jueves 29 de octubre de 2015. Los datos en rojo indican el consumo real, los datos en azul indican las estimaciones.</i>	42
4.6. <i>Predicción del jueves 29 de octubre 2015 con diferentes métodos de estimación. La línea roja representa el consumo real.</i>	44

- 4.7. *Estimaciones para el día jueves de 2015 con la función SE+PER. Los datos en rojo representan el consumo real, los azules las estimaciones* 46
- 4.8. *Estimación semanal de los datos alternos. Los datos en rojo representan el consumo real, los azules las estimaciones. .* 47

LISTA DE TABLAS

	Pág.
2.1. <i>Datos ejemplo de regresión</i>	18
4.1. <i>Resultados de la predicción diaria por semanas con algunas funciones base</i>	38
4.2. <i>Resultados de la predicción diaria por semanas</i>	39
4.3. <i>Resultados de la predicción horaria en el año 2015.</i>	41
4.4. <i>Resultados de las estimaciones con las funciones base.</i>	41
4.5. <i>Media y desviación estándar</i>	42
4.6. <i>Errores métodos alternativos horario entre los años 2013 y 2015</i>	44
. . .	

RESUMEN

El consumo de energía eléctrica se ha incrementado considerablemente en los últimos años debido al crecimiento de la población que cada vez más demanda este tipo de recurso. Se hace indispensable entonces, conocer con antelación y lo más aproximado posible, el consumo de la demanda de energía eléctrica que tendrá cierta población en determinado intervalo de tiempo para así poder vaticinar la cantidad de recursos económicos, humanos y operativos necesarios para generar la cantidad de electricidad necesaria que suplirá dicha demanda. Esta investigación examinó los procesos Gaussianos para abordar el problema del pronóstico de la demanda de energía eléctrica. La metodología propuesta se aplicó para inferir el comportamiento de la demanda de energía usando dos bases de datos reales. Se analizó el rendimiento ante diferentes funciones de covarianza y diferentes métodos del estado del arte también empleados para pronósticos en series de tiempo. Los resultados muestran que la técnica sugerida es una buena alternativa para realizar predicciones.

ABSTRACT

Electricity consumption has increased considerably due to the population has grown in last years. So, it is important to know the amount of consumption in a specific time range in order to prepare necessary resources to supply the demand. This investigation studied the machine learning method known as Gaussian Processes Regression (GPR) for forecasting. The experiments used two real data bases, some basic covariance functions, others prediction methods and one composed covariance function. The results let us see that Gaussian Processes are a good choice for forecasting time series.

INTRODUCCIÓN

1.1. JUSTIFICACIÓN

El consumo de energía eléctrica en Colombia se incrementará considerablemente en los próximos años según estudios y estimaciones de la Unidad de Planeación Minero-Energética (UPME)¹.

Estas proyecciones en los incrementos de demanda de energía eléctrica suponen la necesidad de implementar una metodología que permita predecir en el corto plazo el consumo de este recurso, pues así se puede prever de manera aproximada la cantidad de insumos necesarios para satisfacer la demanda.

En Colombia, la UPME, elabora un pronóstico de la demanda de energía mensual y anual del país. Dicha entidad emplea modelos Autorregresivos como Autoregressive Integrated Moving Average with Explanatory Variable (ARIMAX) [25] y modelo de vectores autorregresivos (VAR) [26]. Este último modelo propone un sistema de ecuaciones con tantas ecuaciones como series a predecir [26].

Entre las críticas que se han hecho a este modelo, se encuentra el número de coeficientes en cada ecuación. En [27] mencionan que, el número de coeficientes en cada ecuación VAR es proporcional al número de variables incluidas en el modelo. Si son demasiados coeficientes, estimarlos

¹<http://www.xm.com.co/Paginas/Consumo/pronostico-de-demanda.aspx>

puede incrementar el error lo cual a su vez puede deteriorar el pronóstico [28]. Otra desventaja es que estos modelos solo son buenos prediciendo en un horizonte corto [29].

También muchos factores que explican las variables se dejan fuera del modelo VAR y entran en el término de error como todo lo demás que explica a la variable [28]. Estos factores pueden estar correlacionados con las variables incluidas, creando problemas de sesgo por omisión de variables [28]. La realización de predicciones en la demanda de energía es un problema que involucra diferentes variables y que pueden estar muy ligadas al éxito en los resultados. Ejemplo de estas variables son el crecimiento económico, el clima, hábitos de consumo, irregularidades temporales tales como días festivos, entre otros. Estos factores pueden causar una discontinuidad en los datos de entrada, por lo tanto, deben ser tratados de manera especial para logra una alta efectividad en los resultados y el modelo en sí. En [30], los autores emplean una meta-función de covarianza que incluye tiempo y temperatura.

¿Por qué procesos Gaussianos?

Aunque existen otras técnicas para realizar predicciones, los procesos Gaussianos se han consolidado como una de las más atractivas para realizar este tipo de tareas.

Una de las características atractivas de los GP es que este modelo reduce la incertidumbre en las zonas cercanas a las observaciones por efectos de la función de covarianza [31]. Esto significa que se puede mejorar el éxito en las estimaciones (reducir el error) si se logra seleccionar una buena función de covarianza.

Otra característica relevante a los GP, es el cálculo de los hiperparámetros, los cuales se pueden obtener analíticamente empleando una de tantas técnicas como lo es Máxima verosimilitud [31]. Es decir, los GP no necesitan la intervención de un experto para obtener los parámetros necesarios para su funcionamiento, como si ocurre con las redes neuronales artificia-

les [33] donde un profesional en el área debe intervenir para configurar su estructura.

Finalmente, la aplicación de los GP para realizar predicciones en otras áreas del conocimiento tales como ingeniería, ciencia, deporte, entre otras, demuestra la versatilidad de esta técnica.

1.2. PLANTEAMIENTO DEL PROBLEMA

La predicción de la demanda es una componente clave para el planeamiento y la operación de los sistemas eléctricos de potencia, donde es usada para la toma de decisiones tales como el despacho, la compra y venta de la electricidad, es también empleada para el análisis de confiabilidad y seguridad de los sistemas de potencia en mercados de energía competitivos [9]. Desde el ámbito de los mercados de energía, esta toma de decisiones está sujeta a factores como el clima, la anticipación para comprar, vender la energía o la alta penetración de la energía renovable, entre otros, introduciendo incertidumbre en la predicción de la demanda de energía y haciendo que esta tarea sea un gran reto de investigación [9].

Con el fin de abordar estos factores mencionados anteriormente, muchos métodos han sido propuestos en la literatura. Estos métodos pueden ser clasificados en dos grupos: los métodos estadísticos y métodos basados en inteligencia artificial [9]. Los enfoques estadísticos usan datos de la demanda de energía para encontrar relaciones lineales entre variables exógenas o periodos de tiempo y la demanda de energía [10, 9]. Entre las técnicas más empleadas dentro de este primer grupo se pueden encontrar: los modelos ARIMA [11]; [12], modelos lineales de regresión [13]; [14], modelos de suavizado exponencial [15], filtros de Kalman [16], entre otros. Estos métodos presentan facilidad en su implementación, sin embargo, el principal inconveniente es el uso de simplificaciones o el hecho de que, por su naturaleza lineal, difícilmente pueden tratar el comportamiento no-lineal de las series de tiempo en el consumo de electricidad. [9]. Por otro lado, los métodos basados en la inteligencia artificial usan grandes cantidades de

datos para realizar la predicción de la demanda de energía. Los métodos más usados son: sistemas expertos [17] máquinas de vectores de soporte [18, 19, 20, 21], redes neuronales [22]. Uno de los inconvenientes de estos métodos inteligentes es depender de la intervención de una persona idónea en el área para la sintonización o especificación de parámetros para realizar una correcta predicción [10]. Es el caso de la Redes Neuronales, donde es necesario que alguien defina la cantidad de capas y neuronas para obtener mejores resultados. Además, los métodos inteligentes no tienen en cuenta la incertidumbre presente en la predicción de la demanda de energía [10].

Además, la característica de no-estacionariedad que se presenta en muchas series de tiempo y que consiste en una variación de la distribución y parámetros a lo largo del tiempo, es tratada por los GP. En [32], los autores emplean una señal de respiración como una serie de tiempo real no estacionaria y emplean una función de covarianza exponencial para realizar predicciones. El experimento se repite, pero con una función de covarianza no-estacionaria, lo cual entrega mejores predicciones.

Con el fin de solucionar los inconvenientes inherentes a las técnicas anteriormente mencionadas (estadísticas y de Inteligencia Artificial), se plantea el uso de los GP como una alternativa para predicción de consumo de energía. Alternativa que no haría parte de ninguna de estas clasificaciones.

Por ejemplo, en [10] han empleado los GP para realizar la predicción de la radiación solar y demanda de energía, mostrando mejor rendimiento en la predicción ante la regresión basada de máquinas de vectores de soporte. Por otro lado, en [23] utilizaron un GP ponderado para incorporar detección de datos espurios para la predicción de un paso adelante de la demanda de energía. Sin embargo, los autores no se enfocan en realizar una predicción con un horizonte un poco más largo, es decir, realizar una predicción de 24 horas adelante. Según Lauret et al. [24], han comparado redes neuronales, redes neuronales Bayesianas y GP para la predicción de la demanda de energía. A partir de este estudio, los autores observan que los GPs superan en rendimiento a las redes neuronales, sin embargo, mencionan que es importante seleccionar la función de covarianza de los GPs,

el cual codifica esa creencia a priori de la forma general de las series de tiempo de la demanda de energía [10].

Con respecto a lo anterior queremos enfocarnos en solucionar las siguientes preguntas de investigación. ¿Será posible encontrar un modelo basado en procesos Gaussianos que sirva para la predicción a corto plazo y que pueda tener en cuenta la no linealidad y la incertidumbre presente en las series de tiempo de la demanda de energía? Además, ¿será posible encontrar o proponer una función de covarianza, la cual pueda ser usada por los procesos Gaussianos para el pronóstico de la demanda de energía?

1.3. ANTECEDENTES

En muchos problemas de regresión, surge la necesidad no sólo de predecir un valor, sino generar también un intervalo de confianza o incertidumbre, puesto que en el problema en cuestión las mediciones realizadas pueden contener ruido o las observaciones pueden verse influenciadas por otras que desconocemos o simplemente no podemos medir [1]. Los GP se consolidan como una alternativa de predicción de series de tiempo que manejan adecuadamente la incertidumbre en el proceso de estimación. Es por ello que en la literatura encontramos una serie de propuestas donde son empleados para este tipo de tareas.

Los autores en [1] realizan un estudio de algunos algoritmos de aprendizaje automático como la regresión lineal, los perceptrones multicapa, las SVR y la regresión lineal Bayesiana, lo cual les ha permitido comprender y concluir el buen rendimiento de los GP.

En [2], el autor se enfoca en una predicción de incertidumbre de demandas por medio de GP y comprara sus resultados con otras dos técnicas, hallando que el GP supera en rendimiento a las otras dos.

En [3] proponen un método probabilístico o exhaustivo basado en procesos gaussianos de regresión cuantil (GPQR). Este método genera una

distribución de probabilidad completa del futuro de la demanda. Se comparan tres diferentes kernels y se selecciona el más adecuado para el modelo, siendo la Exponencial Cuadrática (S.E.) la seleccionada. Los autores emplearon diferentes bases de datos encontrando ventajas tales como que los GPQR no solo ofrecen puntos de pronóstico para la carga futura, sino que también ofrecen una descripción completa de probabilidad para intervalos de predicción, lo cual es especialmente deseable para cuantificar las numerosas incertidumbres de los sistemas de poder [3]. Los resultados de los casos de estudio muestran la calidad de los intervalos de predicción (PI) y la eficiencia de la función de covarianza S.E., pues ésta ayuda a obtener buenos resultados.

En [4], los autores se enfocan en un método híbrido, es decir, un método que reúne algunas de las técnicas más utilizadas en la predicción de la demanda de energía eléctrica. Los métodos empleados son: redes neuronales artificiales (ANN), regresión de soporte de vectores (SVR), máquina de vectores de soporte con mínimos cuadrados (LS-SVM), regresión con procesos gaussianos (GPR) y modelo de mezcla gaussiana (GMM). En este trabajo emplean la función de covarianza S.E. Aunque no es un trabajo que se enfoque solo en los GPR, la inclusión de este método permite suponer que se trata de una de las más adecuadas técnicas para situaciones de predicción de demanda de energía eléctrica, como así lo sugiere la literatura.

En [5], vemos una comparación de dos métodos: proceso gaussiano multitarea y proceso gaussiano monotarea. En problemas de aprendizaje multitarea, la secuencia de datos observados tanto de entrada como de salida de diferentes tareas, están conectadas por variables ocultas. Por lo tanto, información adicional puede ser obtenida desde los datos de las tareas vecinas [5]. Una ventaja de este método es que ante grandes números de observaciones de diferentes tareas se puede presentar una disminución de los ruidos perturbados. De acuerdo a los resultados obtenidos en este trabajo de investigación, el proceso Gaussiano multitarea (Multi-task GP) es el que funciona mejor para predecir la demanda de energía eléctrica.

En [6], encontramos un trabajo donde emplean datos históricos de tres subestaciones de distribución localizadas en Portugal. Acá se considera la temperatura como un factor en el desarrollo del modelo, pues ella está altamente correlacionada con el consumo de energía, particularmente en verano cuando la humedad y la temperatura son altas, o durante el invierno cuando son bajas.

En [7], emplean un conjunto de datos de entrenamiento con consumo histórico de un mes, un año, dos y tres años. Para este estudio emplean GPR y RVR (Relevance Vector Regression), ambos con función de covarianza S.E. Los resultados se presentan en términos de intervalos mensuales dando 12 resultados para los años 2007-2010, y 8 resultados para el año 2011. En esta oportunidad, el GPR no arroja buenos resultados para todos los meses en 2007, excepto para el mes de agosto. Para el año 2008, RVR es el mejor pronosticador en todos los meses, excepto para agosto, donde es ligeramente superado por el GPR.

En [56] encontramos una regresión Gaussiana para modelar una serie de tiempo del consumo de electricidad y flujo de peatones. Los autores emplean el kernel periódico debido a que los datos se enfocan en series de tiempo que exhiben algún grado de periodicidad. El kernel periódico permite el modelamiento de funciones que presentan un patrón repetitivo en su comportamiento a lo largo de un intervalo de tiempo. Por ejemplo, si tenemos unos datos muestreados con una frecuencia horaria y queremos una estructura de repetición diaria en las funciones generadas desde los procesos gaussianos, podemos lograrlo seleccionando un kernel periódico y configurar sus hiperparámetros de periodicidad con un valor de 24.

En [8] tenemos un trabajo donde presentan el pronóstico de la demanda de electricidad entre 2014 a 2024 usando procesos gaussianos el cual usa un conjunto de datos que van desde 2000 a 2013. Para modelar una tendencia a largo plazo, emplean un kernel S.E., para modelar una tendencia media o estacional, usan un kernel que resulta de multiplicar una S.E. con un PER., para modelar una corta y mediana tendencia emplean un kernel Rational Quadratic R.Q.

Según la literatura y los trabajos indicados en esta sección, los GP parecen ser una buena alternativa para realizar predicciones dadas sus características de manejo de incertidumbre, ruido, estacionariedad entre otras, aunque no se puede pasar por alto su elevado costo computacional. Sin embargo, son pocos los trabajos de investigación que emplean funciones de covarianza compuestas, las cuales pueden mejorar las estimaciones, pues al reunir dos o más funciones de covarianza base por medio de la adición o multiplicación, cada una de ellas puede trabajar por separado de acuerdo a sus especialidades o fusionar estas especialidades para tratar de obtener mejores rendimientos.

1.4. OBJETIVOS

Objetivo general

Desarrollar una metodología para la predicción de la demanda de energía eléctrica usando Procesos Gaussianos.

Objetivos específicos

- Proponer una función de covarianza la cual puede ser usada en los Procesos Gaussianos para el pronóstico de la demanda de energía eléctrica.
- Formular un modelo de pronóstico que se base en series temporales de la demanda de energía eléctrica, el cual use Procesos Gaussianos y diferentes funciones de covarianza.
- Validar el método de predicción de la demanda de energía eléctrica y comparar su rendimiento ante enfoques del estado del arte usando diferentes bases de datos y diferentes técnicas de predicción.

Necesidad de pronosticar la demanda de energía

Desde el punto de vista de cada agente en un mercado de electricidad, existen diferentes razones que justifican la necesidad de pronosticar la demanda. Ellas están asociadas a los horizontes de pronóstico utilizados [40]. Para el corto plazo, la predicción se realiza para la hora siguiente y con un horizonte de hasta una semana hacia adelante; el pronóstico se basa en los datos históricos de demanda diaria y factores climáticos. Para el mediano plazo, el pronóstico es mensual con un horizonte de hasta un año. Los datos usados son la demanda histórica de electricidad, y la predicción de diferentes variables explicativas como el crecimiento de la economía o el clima [39]. Para el largo plazo, se abarca un periodo de cinco a veinte años. Tiene un papel importante para prever la necesidad de programar la construcción de nuevas centrales de generación de energía eléctrica, expansión del sistema de transmisión de energía y determinar políticas de regulación de precios [41]. En [42], el autor expone que para pronóstico de largo plazo las variables que más afectan a la demanda de energía son las variables socio-económicas, ya que consideran el crecimiento de la población y del producto interno bruto; por otra parte, las variables climatológicas dejan de influir para este horizonte de tiempo [41].

Para los distribuidores y transmisores, en el corto plazo, permite realizar el monitoreo del comportamiento de la red con el fin de identificar puntos críticos que en el futuro pueden llevar a grandes fallas. Se utiliza, en el mediano plazo, para el planeamiento de los mantenimientos del sistema. Es un insumo fundamental para definir el programa de expansión del sistema de transmisión en el largo plazo [43],[44].

Para los comercializadores, en el mediano y largo plazo, es fundamental para el análisis del comportamiento del mercado y la definición de los planes estratégicos y operativos para la comercialización de la electricidad [43].

Para el operador del mercado, en el corto plazo es necesario para la programación y operación de las unidades de generación, minimizando los

ajustes del programa de operación en tiempo real, con el fin de evitar los sobrecostos en la producción de la electricidad [44].

Características de las series de demanda

Las series de demanda de electricidad presentan características particulares que se han identificado en diferentes estudios. El consumo de electricidad se puede considerar una variable continua en el tiempo y que puede estar agregada en diferentes escalas de tiempo [40],[43],[45]. Se sabe que, para diferentes niveles de agregación temporal, una misma serie puede exhibir complejidades particulares que dificultan el desarrollo de un modelo de predicción [40],[43],[45]. Estas complejidades pueden ser fuertes patrones cíclicos de periodicidad, y comportamientos atípicos como la presencia de días festivos. El consumo de electricidad varía de acuerdo con la estación del año, el día de la semana y la hora del día. Los mayores consumos se presentan durante las estaciones de invierno y verano (en los países que tienen cuatro estaciones), en los días laborales y en las horas pico (estas varían de acuerdo con las características de consumo e industrialización de cada país) [40],[43],[45].

Finalmente, se sabe que existen unas características comunes en las series de demanda de electricidad, las particulares de cada mercado eléctrico hacen que cada serie presente características únicas que la identifican [36].

Serie de demanda anual de electricidad en Colombia

El análisis del consumo anual de energía eléctrica en Colombia (en GWh), comprende el periodo 2003 y 2016. Los datos están disponibles en la página web del operador ².

En la figura 1 podemos observar la tendencia creciente del consumo a lo largo de los 14 años relacionados. Puede observarse un patrón estacional que se repite al final de cada año, es decir, al mes doce.

²<http://informacioninteligente10.xm.com.co/xt>

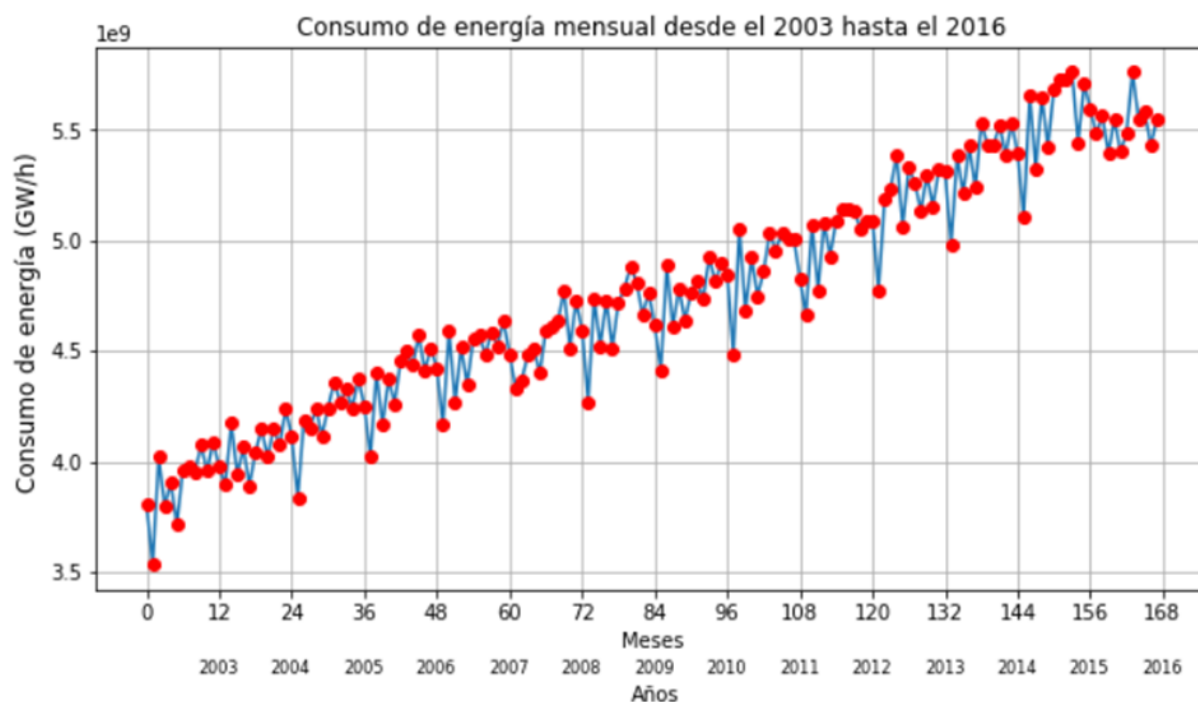


Figura 1.1: *Consumo de energía eléctrica en Colombia entre los años 2003 y 2016.*

Fuente:Elaboración propia.

Para el año 2016 se evidencia un ligero descenso en el consumo. Este comportamiento puede estar relacionado con la aplicación de la resolución CREG 029 de 2016 por la cual se define un esquema de tarifas diferenciales para establecer los costos de prestación del servicio de energía eléctrica a usuarios regulados en el SIN para promover el ahorro voluntario de energía. Dicha resolución fue difundida a nivel nacional a través de la campaña denominada “Ahorrar paga”, impulsada por el gobierno nacional, la cual, derivó en un cambio de hábitos por parte de los usuarios regulados, generando así una disminución en la demanda de energía. Lo anterior impactó el comportamiento de la demanda a partir del mes de marzo de 2016.

PROCESOS GAUSSIANOS

Los procesos Gaussianos son modelos probabilísticos no paramétricos que hacen parte del aprendizaje de máquina como una metodología utilizada en problemas de regresión y clasificación. El aprendizaje consiste en inferir una función a través de un conjunto de datos de entrenamiento [46],[47]. Este conjunto de datos está compuesto por una pareja de valores, una entrada y una salida. La salida es también conocida como valor objetivo y se representa por un escalar, o un vector en el caso de múltiples salidas [46]. El conjunto de datos es utilizado para entrenar el modelo y a partir de este hacer inferencias acerca de nuevas entradas. Un proceso Gaussiano se define como:

Definición: *Un Proceso Gaussiano es una colección de variables aleatorias, posiblemente infinitas las cuales siguen una distribución Gaussiana conjunta*[46].

En otras palabras, significa que la función de distribución de probabilidad conjunta de las variables aleatorias seleccionadas se distribuye normalmente [48]. Un proceso Gaussiano es completamente especificado por su función media $m(x)$ y su función de covarianza $k(x, x')$ [46], denotado como:

$$f(x) \sim GP(m(x), k(x, x')) \quad (2.1)$$

Las variables x y x' , representan los datos de entrenamiento y de prueba respectivamente.

Existe una distinción importante en los procesos Gaussianos, relacionada con la salida del modelo. Cuando la salida es discreta se conocen como Procesos Gaussianos para clasificación, mientras que cuando la salida es continua se conoce como Procesos Gaussianos para Regresión.

2.1. Función media

La función media $m(x)$ corresponde al valor esperado de la función $f(x)$. Generalmente la función media se toma como cero por simplicidad del modelo, aunque no necesariamente debe serlo, según lo propone [46].

$$m(x) = E[f(x)] \quad (2.2)$$

2.2. Función de covarianza

La función de covarianza es también conocida como la función kernel del proceso Gaussiano, y se caracteriza por ser de naturaleza no paramétrica, proporcionando más flexibilidad al modelo. La covarianza entre dos variables aleatorias $f(x)$ y $f(x')$ está dada por la siguiente ecuación [46]:

$$k(x, x') = E[(f(x) - m(x))(f(x') - m(x'))] \quad (2.3)$$

La ecuación anterior puede simplificarse si se toma la función media como cero, esto es $m(x) = 0$, $m(x') = 0$, se escribe entonces la función de covarianza como:

$$k(x, x') = E[(f(x)f(x'))] \quad (2.4)$$

El kernel es la función más significativa en los modelos de aprendizaje basados en procesos Gaussianos, puesto que la precisión del modelo se debe principalmente a la función de covarianza seleccionada [48]. Para la selección de la función kernel se deben tener en cuenta ciertas características según lo propone, Williams y Rasmussen [46]. Una de ellas es que la función seleccionada debe ser definida semi-positiva y simétrica, es decir $k(x, x') = k(x', x)$.

Dos de las funciones de covarianza base más comunes son la Exponencial cuadrática o Squared Exponenetial (SE) y la Periódica (PER), las cuales se detallan a continuación.

2.2.1. Exponencial Cuadrática

Esta función también es conocida como Radial Basis Function (RBF) y se define como:

$$SE(x, x') = \sigma_{SE}^2 \exp\left(-\frac{(x - x')^2}{2\lambda_{SE}^2}\right) \quad (2.5)$$

Posee dos parámetros, λ_{SE} y σ_{SE}^2 . El primero se conoce como la longitud de escala, el cual controla la escala horizontal sobre la cual la función cambia. El segundo es la varianza del ruido el cual controla los cambios en la escala vertical [49]. Esta función es ampliamente usada en los procesos

gaussianos por su capacidad de lizado o smoothing [50].

2.2.2. Periódica

Una función de covarianza periódica se define como:

$$PER(x, x') = \sigma_{PER}^2 \exp\left(-\frac{2 \sin^2(\pi(x - x')/p)}{\lambda_{PER}^2}\right) \quad (2.6)$$

Un kernel periódico ofrece una estructura repetitiva sobre la función, la cual es controlada por el parámetro del periodo p . Un valor de p demasiado grande causará oscilaciones bajas, mientras que un valor pequeño causará oscilaciones altas [51]. La longitud de escala λ_{PER} y σ_{PER}^2 tienen el mismo efecto que se mencionó para el kernel SE.

2.3. El Conjunto de datos de entrenamiento

El conjunto de datos de entrenamiento D está formado por un conjunto de parejas (x, y) donde x representa el vector de entradas y y es un escalar que corresponde a la salida o valor objetivo. El conjunto de datos de entrenamiento para n observaciones se escribe entonces:

$$D = [(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)] = [(x_i, y_i)_{i=1,2,\dots,n}]$$

Donde n es la cantidad de parejas de datos.

Si se considera ahora x un vector columna de dimensión D , es decir, $x = (x_1, x_2, \dots, x_D)^\top$ es posible agrupar el conjunto de datos de entrada en la matriz X de dimensiones D x n . Las correspondientes salidas son indexadas en el vector y , esto es, $y = (y_1, y_2, \dots, y_n)^\top$. Así, el conjunto de datos de entrenamiento se escribe de la siguiente manera [46]:

$$D = (X, y)$$

2.4. Análisis de regresión

El análisis de regresión en el aprendizaje supervisado, parte de un conjunto de muestras D , y tiene como objetivo inferir una función capaz de predecir el valor de la salida y_* y el error asociado correspondiente a una nueva entrada x_* que no pertenece al conjunto de entrenamiento.

Para ilustrar mejor la idea anterior considere el conjunto de datos que se muestran en la tabla 2.1 conformado por cinco valores de entrada y su correspondiente salida. El objetivo de la regresión es predecir el valor de salida para la nueva entrada $x = 6$ y su respectivo error. La figura 2.1 muestra el comportamiento de los datos. Las barras asociadas a cada valor representan el ruido o el error de cada observación.

Tabla 2.1: *Datos ejemplo de regresión*

x	1	2	3	4	5	6
y	0.5	0.7	0.9	1.3	1.7	?

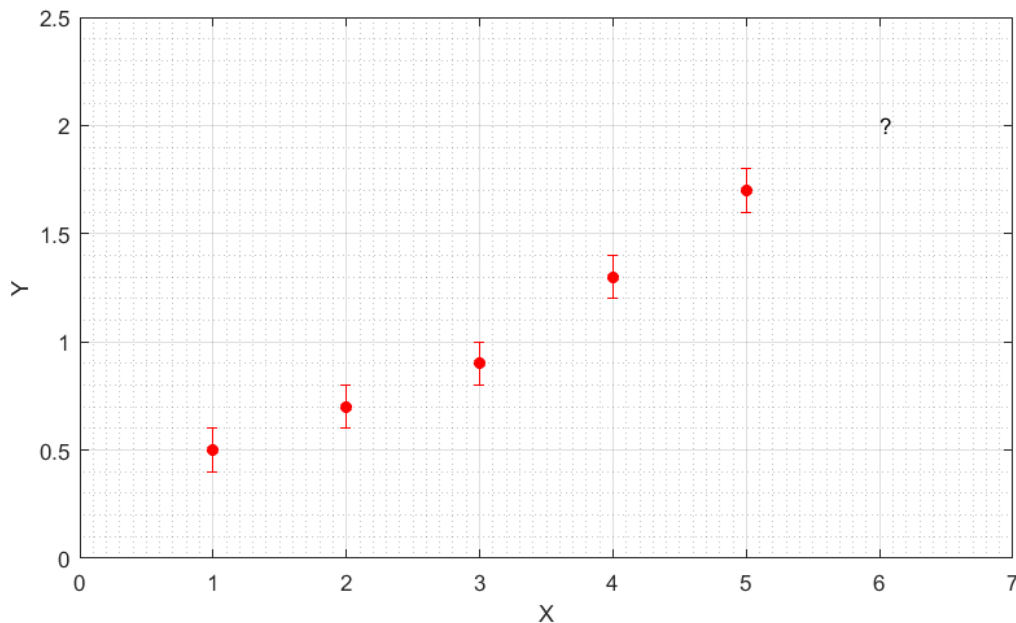


Figura 2.1: *Ejemplo datos de entrenamiento. La barra representa el ruido de cada observación*

Fuente:Elaboración propia.

2.5. Procesos gaussianos para regresión

La regresión con procesos gaussianos parte de un conjunto de entrenamiento D conformado por n observaciones. El objetivo es aprender una función para predecir el valor objetivo y_* correspondiente a una nueva entrada x_* . Normalmente no se tiene conocimiento de los valores reales de la función, la información disponible se encuentra con ruido (ϵ) el cual se supone que sigue una distribución gaussiana, $(\epsilon) \sim N(0, \sigma_n^2)$ con media 0 y varianza σ_n^2 .

El valor objetivo puede representarse mediante el modelo de regresión lineal estándar con ruido aditivo, como se muestra a continuación:

$$y = f(x) + \epsilon \quad (2.7)$$

La función media a priori se asume como $x = 0$, es decir se supone que la distribución apriori de las funciones se distribuyen normalmente con media cero; de esta manera es posible simplificar la función de covarianza, aunque como se mencionó anteriormente, puede ser diferente de cero. La función de covarianza $cov(x, x')$ para las observaciones ruidosas se calculan como:

$$cov(x, x') = k(x, x') + \sigma_n^2 \delta(x, x') \quad (2.8)$$

Donde $\delta(x, x')$ es la función delta de Kronecker, la cual es igual a 1 si $x = x'$ o cero de cualquier otra forma. La función de covarianza se evalúa para encontrar la correlación entre los datos del conjunto de entrenamiento, entre los datos del conjunto de prueba y los datos de entrenamiento y entre los datos de prueba, recordando que estos se asumen portadores de un ruido gaussiano [52].

2.6. Matriz de covarianza

2.6.1. Covarianza del conjunto de datos de entrenamiento

Al evaluar la función de covarianza entre cada uno de los puntos de los datos del conjunto de entrenamiento formado por n observaciones, se obtiene la matriz de covarianza K de dimensión $n \times n$.

$$K = \begin{bmatrix} cov(x_1, x_1) & cov(x_1, x_2) & \dots & cov(x_1, x_n) \\ cov(x_2, x_1) & cov(x_2, x_2) & \dots & cov(x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ cov(x_n, x_1) & cov(x_n, x_2) & \dots & cov(x_n, x_n) \end{bmatrix} \quad (2.9)$$

2.6.2. Covarianza del conjunto de datos de prueba

La matrix de covarianza entre los datos de prueba y ellos mismos viene representada por K_{**} y se escribe como:

$$K_{**} = cov(x_*, x_*) \quad (2.10)$$

2.6.3. Covarianza del conjunto de entrenamiento y conjunto de prueba

La correspondiente matrix de covarianza entre los datos de entrenamiento y los datos de prueba está dada como K_* y se representa como

$$K_* = \begin{bmatrix} cov(x_*, x_1) & cov(x_*, x_2) & \dots & cov(x_*, x_n) \end{bmatrix} \quad (2.11)$$

La distribución conjunta de observaciones y y de las predicciones y_* es obtenida como una distribución gaussiana multivariante como se muestra a continuación [46, 48]:

$$\begin{bmatrix} y \\ y_* \end{bmatrix} \sim N \left(0, \begin{bmatrix} K & K_*^T \\ K_* & K_{**} \end{bmatrix} \right) \quad (2.12)$$

Ahora se establece la distribución condicional de las distribuciones de los datos de prueba dadas las distribuciones de los datos de entrenamiento, esto es $p(y_*|y)$, y se expresa como lo indica la ecuación 2.13. Esta distribución se conoce como la distribución a posteriori. Para una distribución gaussiana la media de la distribución a posteriori se conoce como el *máximo a posteriori* (MAP), el cual es la mejor estimación de la variable analizada [46, 48].

$$y_*|y \sim N(K_*K^{-1}y, K_{**} - K_*K^{-1}K_*^T) \quad (2.13)$$

La estimación MAP de la distribución anterior está dada por su media, el cual representa el mejor valor esperado de la predicción del Proceso Gaussiano. Mientras el error esta dado por la varianza, como se muestra a continuación: [46, 48].

$$\bar{y}_* = K_*K^{-1}y \quad (2.14)$$

La varianza de la estimación está dada por

$$var(y_*) = K_{**} - K_*K^{-1}K_*^T \quad (2.15)$$

Las expresiones 2.14 y 2.15 conforman las ecuaciones fundamentales de un proceso gaussiano para regresión.

2.7. Hiperparámetros

La funciones de covarianza como se mencionó anteriormente es la función principal del Proceso Gaussiano, determina la forma en que se relacionan los datos, determina la efectividad del modelo y la precisión de las predicciones en el análisis de regresión. La función de covarianza seleccionada está compuesta por un conjunto de parámetros libres, definidos sobre la distribución a priori y reciben el nombre de *Hiperparámetros* los cuales pueden ser valores constantes o distribuciones de probabilidad [46, 47].

2.7.1. Determinación de los hiperparámetros

En procesos Gaussianos para regresión los parámetros se obtienen mediante aprendizaje de los datos de entrenamiento. Existen diferentes técnicas para la estimación de estos valores, entre estos el modelo de inferencia Bayesianos conocido como *Maximización de la probabilidad marginal*. El método consiste en maximizar la probabilidad marginal $p(y|X, \theta)$ [48], donde

$$\theta = [l, \sigma_f^2, \sigma_n^2]$$

Siendo l la longitud de escala, σ_f^2 la varianza de la señal y σ_n^2 la varianza del ruido.

La probabilidad a posteriori de los parámetros puede escribirse aplicando el teorema de Bayes como se muestra a continuación [48].

$$p(\theta|X, y) = \frac{p(y|X, \theta)p(\theta)}{p(y|X)} \quad (2.16)$$

donde

- $p(\theta|X, y)$ es la probabilidad a posteriori de los parámetros
- $p(y|X, \theta)$ es la probabilidad marginal, función a maximizar
- $p(\theta)$ es la probabilidad apriori de los parámetros
- $p(y|X)$ es la probabilidad de las salidas dadas las entradas. Probabilidad apriori de los datos o verosimilitud marginal.

La probabilidad marginal puede ser calculada por marginación de la integral de probabilidad y la probabilidad a priori sobre los valores de la función latente f , libres de ruido [48].

$$p(y|X, \theta) = \int p(y|f, X)p(f|X)df \quad (2.17)$$

Ambas $p(y|f, X)$ y $p(f|X)$ siguen distribuciones gaussianas.

Se calcula el valor del logaritmo de la probabilidad marginal de los parámetros como [48],[46]:

$$L(\theta) = \log[p(y|X, \theta)] = -\frac{1}{2}y^T K^{-1}y - \frac{1}{2}\log|K| - \frac{n}{2}\log 2\pi \quad (2.18)$$

En la anterior ecuación se distinguen tres términos. El primero $-\frac{1}{2}y^T K^{-1}y$ representa el ajuste de los datos e involucra el valor objetivo de las observaciones. El segundo término $-\frac{1}{2}\log|K|$ representa el valor de penalización complejo. Y el tercero $-\frac{n}{2}\log 2\pi$ es una constante de normalización [46].

El conjunto de hiperparámetros se determinan maximizando la ecuación anterior. Para ello es necesario calcular las derivadas parciales de la

probabilidad marginal con respecto a cada uno de los parámetros.

$$\begin{aligned}\frac{\partial}{\partial \theta_j} \log p(y|X, \theta) &= \frac{1}{2} y^T K^{-1} \frac{\partial K}{\partial \theta_j} K^{-1} y - \frac{1}{2} \text{tr}(K^{-1} \frac{\partial K}{\partial \theta_j}), \\ &= \frac{1}{2} \text{tr}[(\alpha \alpha^T - K^{-1}) \frac{\partial K}{\partial \theta_j}],\end{aligned}\tag{2.19}$$

donde $\alpha = K^{-1}y$ y tr es la traza de una matrix. Una vez calculada las derivadas parciales, se puede aplicar cualquier método de optimización basado en el gradiente [46].

2.8. Composición de funciones de covarianza

Es posible combinar algunas funciones de covarianza base para modelar una función más compleja. Es primordial que la matrix de covarianza resultante sea definida semi-positiva. La adición y la multiplicación son dos operaciones que permiten crear una nueva función de covarianza [46].

Esta nueva función de covarianza se define como $\hat{k}(x, x')$. Y el proceso gaussiano como

$$f(t) \sim GP(0, \hat{k}(x, x')),$$

donde la función de covarianza $\hat{k}(x, x')$ es una combinación de kernels por medio de la adición y multiplicación. Por ejemplo, la función de covarianza podría tener la forma

$$\hat{k}(x, x') = k_1(x, x') + k_2(x, x').k_3(x, x') + \dots\tag{2.20}$$

donde $k_i(x, x')$, $i = 1, 2, 3...$ puede ser una función de base radial (RBF), periódica (PER), Lineal (LIN) o cualquier otra función de covarianza base. La idea al realizar una composición de una función de covarianza es que cada función base sea responsable de modelar una cierta estructura o patrón en la series de tiempo de acuerdo a sus características [51].

El modelo por adición establece que una serie de tiempo puede ser definida por su tendencia, temporalidad y componentes residuales [51].

$$y_t = t_t + s_t + e_t \quad (2.21)$$

donde t_t , s_t y e_t denotan la tendencia, temporalidad y componente residual respectivamente. Una serie de tiempo contiene una tendencia cuando los datos exhiben un comportamiento similar a alguna función matemática. Esta tendencia puede ser lineal, cuadrática o cualquier otro comportamiento no-lineal. Un patrón temporal es una secuencia repetitiva de los datos que en la mayoría de los casos no es exacta. El componente residual es aquella parte del modelo que no puede ser explicado ni por la tendencia ni por la temporalidad [51].

Por ejemplo, una tendencia lineal puede ser modelada con el kernel lineal, mientras que la temporalidad puede ser modelada con el kernel periódico. Para este caso, tendríamos una función compuesta

$$\hat{k}(x, x') = k_{LIN}(x, x') + k_{PER}(x, x') \quad (2.22)$$

De manera similar, si necesitamos modelar una tendencia cuadrática se puede multiplicar dos funciones base lineales (LINxLIN). Si el componente temporal no sigue unas oscilaciones exactas se puede, además de la función base periódica (PER), agregar la función RBF. La nueva función de

covarianza sería

$$\hat{k}(x, x') = k_{LIN}(x, x') \cdot k_{LIN}(x, x') + \hat{k}_{PER}(x, x') + k_{RBF}(x, x') \quad (2.23)$$

Algunas arquitecturas, tales como redes convolucionales [53] y sum-product networks[54], incluyen unidades que computan operaciones de tipo AND y OR. La composición de kernels puede ser comprendida de esta manera [55]. Una suma de kernels se puede entender como una operación OR. De manera similar, la multiplicación puede verse como una operación AND [56].

En la adición, las características de cada kernel trabajan de manera independiente, mientras que en la multiplicación las características de los kernel se fusionan para trabajar de manera unificada [51].

MATERIALES Y MÉTODOS

En esta sección se presentan los materiales y métodos usados para el pronóstico de la demanda de energía usando GP. Primero, se explica las dos bases de datos empleadas. Seguidamente, se presenta la forma en la que se depuraron los datos para su posterior uso en las etapas de entrenamiento y validación usando los procesos Gaussianos.

El tratamiento previo de los datos fue realizado en MICROSOFT® EXCEL. Los algoritmos de inferencia del GP fueron programados en MATLAB®, utilizando el toolbox GPML creado por Carl Edward Rasmussen y Hannes Nickisch[57].

3.1. Origen de los datos

El conjunto de datos empleados en este trabajo fue obtenido desde la página web de la empresa XM¹, para el caso de consumo en Colombia, y de Belgian Electricity Supplier Elia² como datos alternativos para comparación y prueba.

XM es una Compañía de Expertos en Mercados, es una filial de la firma estatal colombiana de transmisión ISA. Presta servicios de administración y gestión de sistemas transaccionales y plataformas tecnológicas en tiem-

¹<http://informacioninteligente10.xm.com.co>

²<http://www.elia.be/>

po real. En la página web de mencionada empresa podemos encontrar el consumo de energía eléctrica comercial por comercializador, desde el año 2000 hasta el 2016 dividido por semestres. Para esta propuesta, se emplea el consumo del año 2015.

3.2. Depurado de los datos

Lo adecuado para realizar las predicciones, es que los datos de entrenamiento sigan un comportamiento regular o similar en un intervalo de tiempo, pues cuando presentan irregularidades, esto afecta de manera negativa la precisión de las estimaciones. Por esta razón, en esta investigación se trabajan dos horizontes a saber: diario y horario.

3.2.1. Diario

Para este horizonte se toman 47 semanas en total, de las cuales 38 serán para train y 9 para test. Se inicia con cuatro semanas de entrada y la siguiente como salida. Luego se avanza una semana y se repite el proceso.

El conjunto de datos de entrenamiento $D = (X, y)$ se define como:

$$X = \begin{bmatrix} S1 & S2 & S3 & S4 \\ S2 & S3 & S4 & S5 \\ S3 & S4 & S5 & S6 \\ \vdots & \vdots & \vdots & \vdots \\ S_{n-3} & S_{n-2} & S_{n-1} & S_n \end{bmatrix} y = \begin{bmatrix} S5 \\ S6 \\ S7 \\ \vdots \\ S_{n+1} \end{bmatrix}$$

Donde $n = (38 \text{ semanas} \times 7 \text{ días}) = 266$.

Y el conjunto de datos de prueba $D_* = (X_*, y_*)$ como:

$$X_* = \begin{bmatrix} S1_* & S2_* & S3_* & S4_* \\ S2_* & S3_* & S4_* & S5_* \\ \vdots & \vdots & \vdots & \vdots \\ S_{*m-3} & S_{*m-2} & S_{*m-1} & S_{*m} \end{bmatrix} y = \begin{bmatrix} S5_* \\ S6_* \\ \vdots \\ S_{*m+1} \end{bmatrix}$$

Donde $m=(9 \text{ semanas} \times 7 \text{ días})=63$

3.2.2. Horario

Para la predicción horaria se sigue la misma mecánica de la predicción diaria. Es decir, se toman los cuatro días del mes de enero y se tiene como valor objetivo el mismo día del siguiente mes. Se avanza un día hacia adelante y se repite el proceso. Este procedimiento se aplica para los días hábiles. En total se toman 48 días de los cuales 34 son para train y 14 para test.

La matriz de los datos de entrenamiento $D = (X, y)$ queda definida así:

$$X = \begin{bmatrix} D1 & D2 & D3 & D4 \\ D2 & D3 & D4 & D5 \\ D3 & D4 & D5 & D6 \\ \vdots & \vdots & \vdots & \vdots \\ D_{n-3} & D_{n-2} & D_{n-1} & D_n \end{bmatrix} y = \begin{bmatrix} D5 \\ D6 \\ D7 \\ \vdots \\ D_{n+1} \end{bmatrix}$$

Donde $n=(34 \text{ días} \times 24 \text{ horas})=816$.

Y el conjunto de datos de prueba $D_* = (X_*, y_*)$ como:

$$X_* = \begin{bmatrix} D1_* & D2_* & D3_* & D4_* \\ D2_* & D3_* & D4_* & D5_* \\ D3_* & D4_* & D5_* & D6_* \\ \vdots & \vdots & \vdots & \vdots \\ D_{*m-3} & D_{*m-2} & D_{*m-1} & D_{*m} \end{bmatrix} y = \begin{bmatrix} D5_* \\ D6_* \\ D7_* \\ \vdots \\ D_{*m+1} \end{bmatrix}$$

Donde $m=(14 \text{ días} \times 24 \text{ horas})=336$.

3.3. Métricas de rendimiento

El desempeño de los modelos implementados en este trabajo es medido mediante el cálculo de dos medidas de error asociados a los resultados obtenidos. Estas dos medidas son el error cuadrático medio estandarizado (SMSE) y el Mean Standardized Log Loss (MSLL), los cuales se describen a continuación.

3.3.1. Error medio cuadrático estandarizado

El error medio cuadrático estandarizado es obtenido mediante normalización del error medio cuadrático (MSE), el cual describe el promedio de la diferencia entre los puntos de validación y_{*i} y la media de las predicciones $\bar{f}(x_{*i})$, como se observa en la siguiente ecuación [47].

$$MSE = (y_{*i} - \bar{f}(x_{*i}))^2. \quad (3.1)$$

La ecuación anterior es normalizada haciendo uso del conjunto de datos de validación, obteniendo la siguiente expresión [47]:

$$SMSE = \frac{1}{\sigma_{test}^2 + \mu_{test}^2} (y_{*i} - \bar{f}(x_{*i}))^2, \quad (3.2)$$

donde (x_{*i}, y_{*i}) son los valores de una observación entrada y salida del conjunto de datos de prueba, σ_{test}^2 y μ_{test}^2 son la varianza y la media del conjunto de datos de prueba respectivamente [47].

3.3.2. Media estandarizada del logaritmo de las perdidas

Para incluir la varianza en el cálculo del error, se toma el logaritmo de la verosimilitud marginal de y_* dados los datos de entrenamiento y validación y el hecho de haber asumido un ruido Gaussiano. Se emplea el negativo del logaritmo de tal forma que se obtengan valores más negativos cuando el modelo es mejor [47].

El MSLL se calcula mediante la siguiente expresión:

$$MSLL = \frac{1}{2} \log(\sigma_{*i}^2) + \frac{(y_{*i} - \bar{f}(x_{*i}))^2}{\sigma_{*i}^2} - \frac{(\sigma_{test}^2 + \mu_{test}^2)}{2}, \quad (3.3)$$

donde $\bar{f}(x_{*i})$ y σ_{*i}^2 son la media y la varianza predictiva en el punto de prueba x_{*i}

3.4. Función de covarianza propuesta

Inicialmente, al hacer el depurado de los datos se observó que éstos presentaban un comportamiento periódico y algunas crestas. Esta es la razón por la cual se involucran las funciones base Exponencial cuadrática(SE) y Periódica (PER), pues como se indicó en la sección 2.2.1 y 2.2.2, estas dos funciones son adecuadas para procesar este tipo de comportamiento. Con el ánimo de mejorar los resultados, se agregaron otras funciones base por adición y multiplicación a esta “meta-función” empleando la técnica del árbol de búsqueda [51] pero no se obtuvo mejores efectos. Finalmente, se

decide emplear el operador de adición pues según la literatura [56] cuando se emplea este operador cada función involucrada trabaja de manera independiente con sus propias características [58].

Entonces, de acuerdo a las razones anteriormente expuestas y a lo descrito en la sección 2.8, la función de covarianza elegida para la elaboración de esta propuesta es **SE+PER**. La expresión más detallada de dicha función es

$$\hat{k}(x, x') = \sigma_{SE}^2 \exp\left(-\frac{(x - x')^2}{2\lambda_{SE}^2}\right) + \sigma_{PER}^2 \exp\left(-\frac{2\sin^2(\pi(x - x')/p)}{\lambda_{PER}^2}\right), \quad (3.4)$$

donde los parámetros λ, σ^2 y p son descritos en las secciones 2.2.1 y 2.2.2

3.5. Otros métodos de predicción

Se presentan otros métodos empleados para realizar estimaciones y su comparación con los procesos gaussianos.

Máquinas de vectores soporte

(SVM, del inglés Support Vector Machines) tienen su origen en los trabajos sobre la teoría del aprendizaje estadístico y fueron introducidas en los años 90 por Vladimir Vapnik y sus colaboradores [59]. Fueron, inicialmente usadas para clasificación, pero su uso fue extendido a tareas de regresión [60]. Una máquina de vectores de soporte construye un hiperplano óptimo en forma de superficie de decisión, de modo que el margen de separación entre las dos clases en los datos se amplía al máximo. Los vectores de soporte hacen referencia a un pequeño subconjunto de las observaciones de entrenamiento que se utilizan como soporte para la ubicación óptima de la superficie de decisión³. En el caso de que las observaciones no puedan

³<https://la.mathworks.com/discovery/svm-maquina-vectores-soporte.html>

ajustarse por una función lineal, se recurre a una metodología similar a la utilizada en el problema de clasificación para datos no separables linealmente. Es decir, los ejemplos pertenecientes al espacio original de entradas se transforman en un nuevo espacio, denominado también espacio de características, en el que sí es posible ajustar los ejemplos transformados mediante un regresor lineal [61]. El tipo de transformación dependerá del kernel utilizado. Para este estudio se emplea un kernel de función de base radial (RBF) o Gaussiano. Para mayor información de las máquinas de vectores de soporte, ver [60],[62],y[63].

Modelo ARIMA

Box y Jenkins han desarrollado modelos estadísticos para series temporales que tienen en cuenta la dependencia existente entre los datos, esto es, cada observación en un momento dado es modelada en función de los valores anteriores. Los análisis se basan en un modelo explícito. Estos modelos se conocen con el nombre de AutoRegressive Integrated Moving Average, que deriva de sus tres componentes: AR (autorregresivo), establece la relación entre un valor determinado y otro anterior (laggeado). I (integrado), se refiere a la capacidad de diferenciar la serie para eliminar la tendencia y la variabilidad creciente o decreciente. MA (medias móviles) representa el error del modelo como combinación de términos de error anteriores.

El modelo ARIMA permite describir un valor como una función lineal de datos anteriores y errores debidos al azar, además, puede incluir un componente cíclico o estacional. Es decir, debe contener todos los elementos necesarios para describir el fenómeno.

Redes Neuronales Artificiales

Artificial Neural Network(ANN). Fueron desarrolladas a partir de los trabajos de investigación sobre la fisiología de las neuronas biológicas de Erlanger y Gasser en el año 1924, y de los estudios el año 1956. Una red

neuronal artificial emula una red neuronal sobre la actividad de los neurotransmisores de Hadgkin y Huxley en biológica en la que cada nodo o unidad de procesamiento, se corresponde con una neurona y tiene como función recibir varios impulsos o entradas, procesarlos y transmitir un resultado a otros nodos o salidas. La implementación de las redes neuronales artificiales como modelo matemático para las redes neuronales biológicas fue desarrollada por McCulloch y Pitts en el año 1943. Sin embargo, como entidad matemática, ellas tienen un interés intrínseco y su importancia y aplicación se ha extendido más allá de lo originalmente imaginado. Las redes neuronales artificiales han resultado apropiadas para el análisis de datos generados en una amplia variedad de disciplinas. En este sentido, las redes no sólo han sido empleadas de forma innovativa e imaginativa para analizar grandes y complejas bases de datos, sino que también han sido utilizadas para resolver problemas tradicionalmente ligados al análisis estadístico. La aplicación de las redes neuronales artificiales a la predicción con series de tiempo no es nueva. Existen numerosos trabajos al respecto, siendo probablemente los más conocidos los de Werbos en los años 1974 y 1988, Lapedes el año 1987, Weigend y sus colegas el año 1990, entre otros.

Redes Neuronales Profundas

Deep Neural Network (DNN). Una red neuronal profunda es una Red Neuronal Artificial (ANN) con múltiples capas ocultas entre las capas de entrada y salida. De forma similar a las Redes Neuronales superficiales, la DNN pueden modelar relaciones complejas no lineales. En [66] y [67] podemos encontrar algunos trabajos donde se emplean las DNN para predicción de consumo de energía eléctrica en un corto plazo.

RESULTADOS EXPERIMENTALES

En este capítulo se presentan los resultados obtenidos en las diferentes pruebas realizadas con los datos y horizontes planteados. Los rendimientos fueron medidos con las métricas mencionadas en el capítulo anterior.

4.1. Resultados predicción diaria

A continuación, se relacionan los resultados de las diferentes estimaciones por semana del año 2015 usando la función de covarianza compuesta SE+PER.

La figura 4.1 permite apreciar gráficamente la estimación para la semana 44 del año 2015, la cual va desde el 26 de octubre hasta el 1 de noviembre, y cuyo error SMSE fue de 0.0001

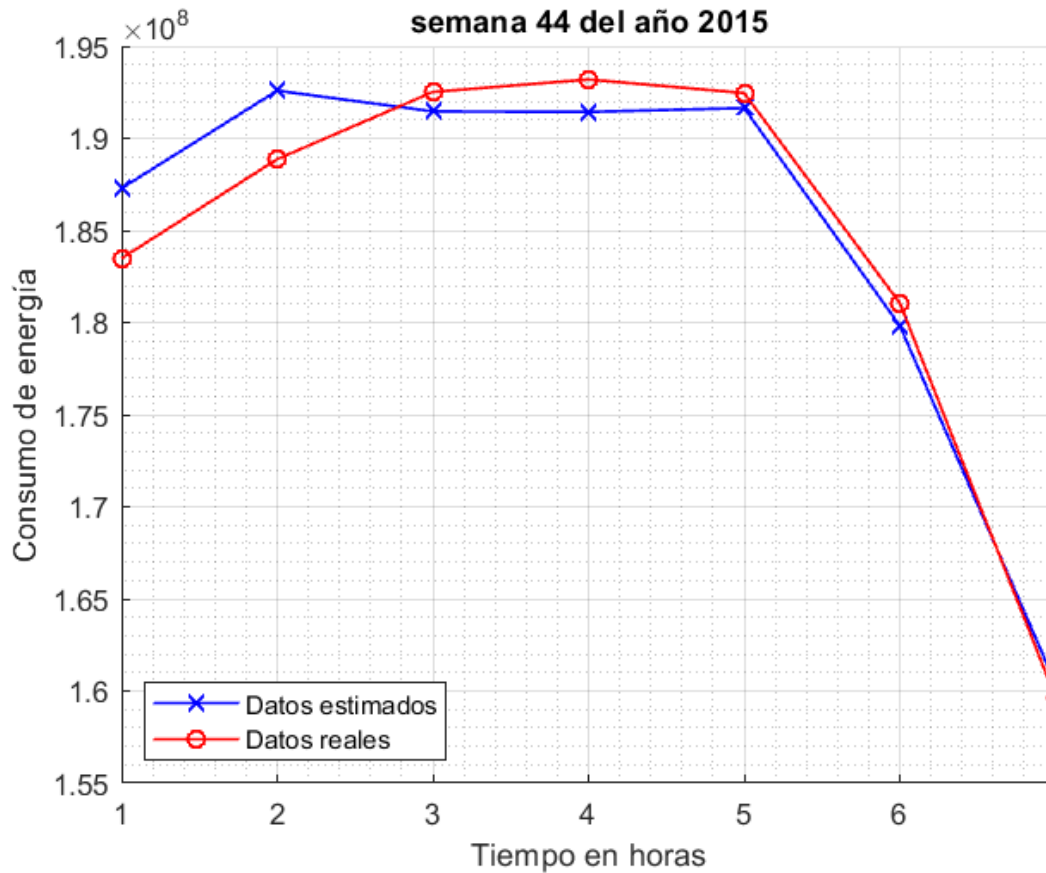


Figura 4.1: Estimación de la semana 44 del año 2015. La línea roja representa el consumo real. La línea azul representa las estimaciones.

Fuente: Elaboración propia.

Con el fin de contrastar la función de covarianza compuesta contra algunas funciones base, el experimento se repite con algunas de ellas a saber: SE, PER y RQ. Los resultados SMSE para la semana 44 del año 2015 con estas funciones base son 0.0002, 0.0036 y 0.0002 y pueden apreciarse visualmente en las gráficas 4.2, 4.3 y 4.4 respectivamente. La tabla 4.1 permite apreciar de manera general los resultados de la repetición del experimento con las funciones mencionadas.

Tabla 4.1: Resultados de la predicción diaria por semanas con algunas funciones base

Función	SMSE	MSLL
SE	0.002562 ± 0.003372	-1.6882 ± 8.9359
PER	0.004231 ± 0.001985	-1.6894 ± 8.7413
RQ	0.002435 ± 0.003380	-1.6885 ± 8.9167

SMSE es el error cuadrático medio estandarizado. Ambas métricas se definen entre los valores reales y los estimados.

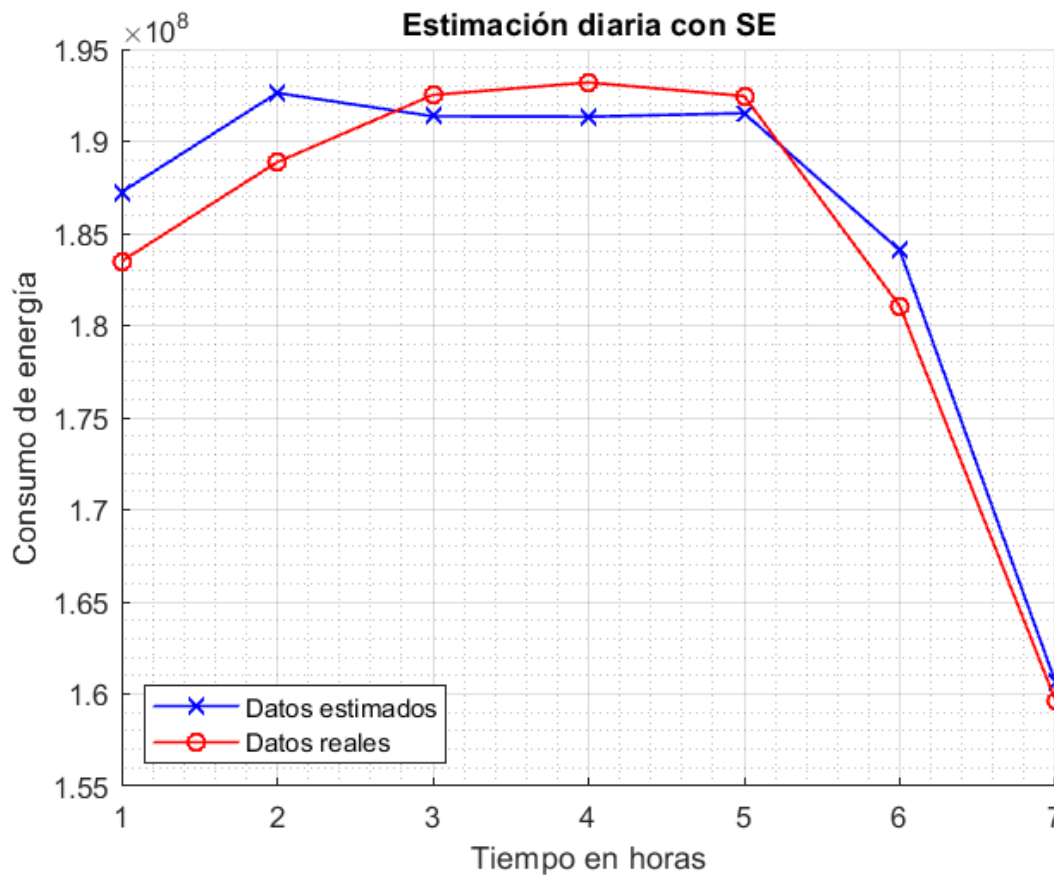


Figura 4.2: Estimación diaria con la función de covarianza base SE. La línea roja representa el consumo real. La línea azul representa las estimaciones.

Fuente: Elaboración propia.

La tabla 4.2 nos permite observar el rendimiento del modelo al repetir el experimento en los años 2014 y 2013. Dicha tabla informa que la tendencia del rendimiento se mantiene estable, lo que indica que el modelo se

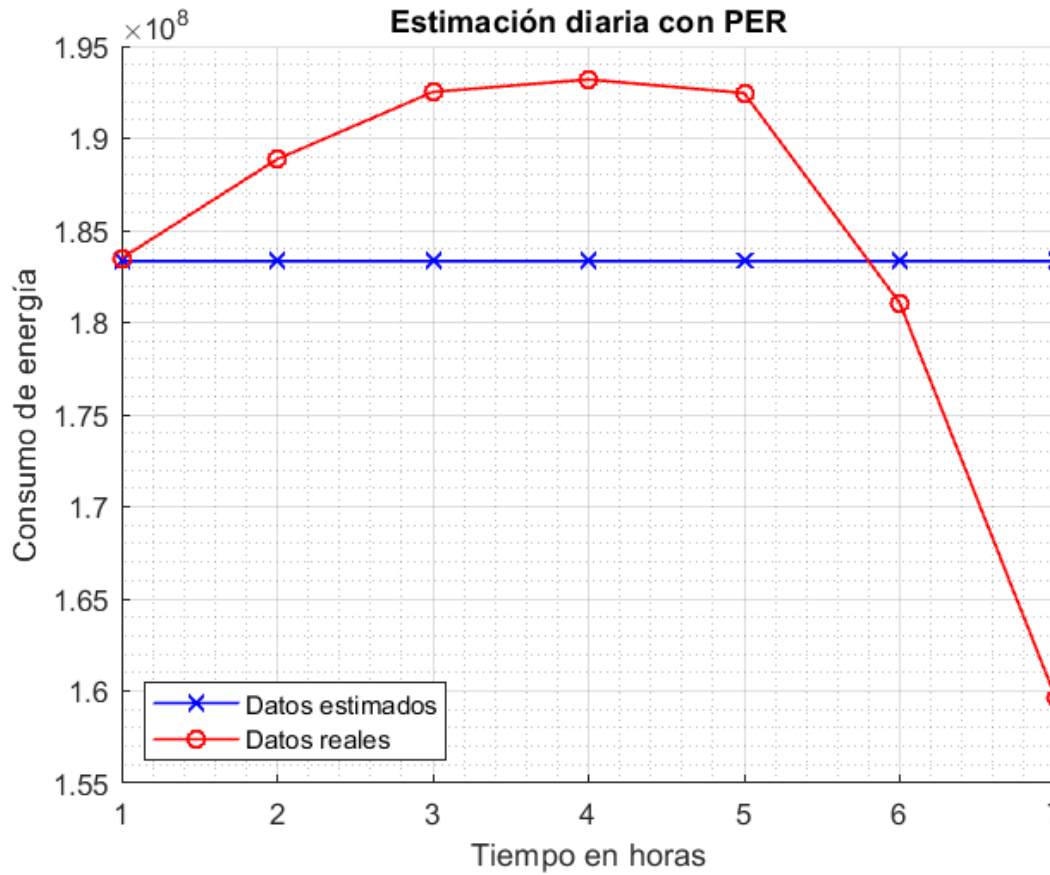


Figura 4.3: *Estimación diaria con la función de covarianza base PER. La línea roja representa el consumo real. La línea azul representa las estimaciones.*

Fuente: Elaboración propia.

comporta de manera similar ante los diferentes datos.

Tabla 4.2: *Resultados de la predicción diaria por semanas*

Año	SMSE	MSLL
2015	0.002372 ± 0.002634	-1.6883 ± 8.8903
2014	0.002946 ± 0.003662	-1.6206 ± 1.0773
2013	0.002761 ± 0.002992	-1.4923 ± 8.2823

SMSE es el error cuadrático medio estandarizado. Ambas métricas se definen entre los valores reales y los estimados.

Al comparar los resultados de la tabla 4.1 con los de la 4.2(en el 2015),

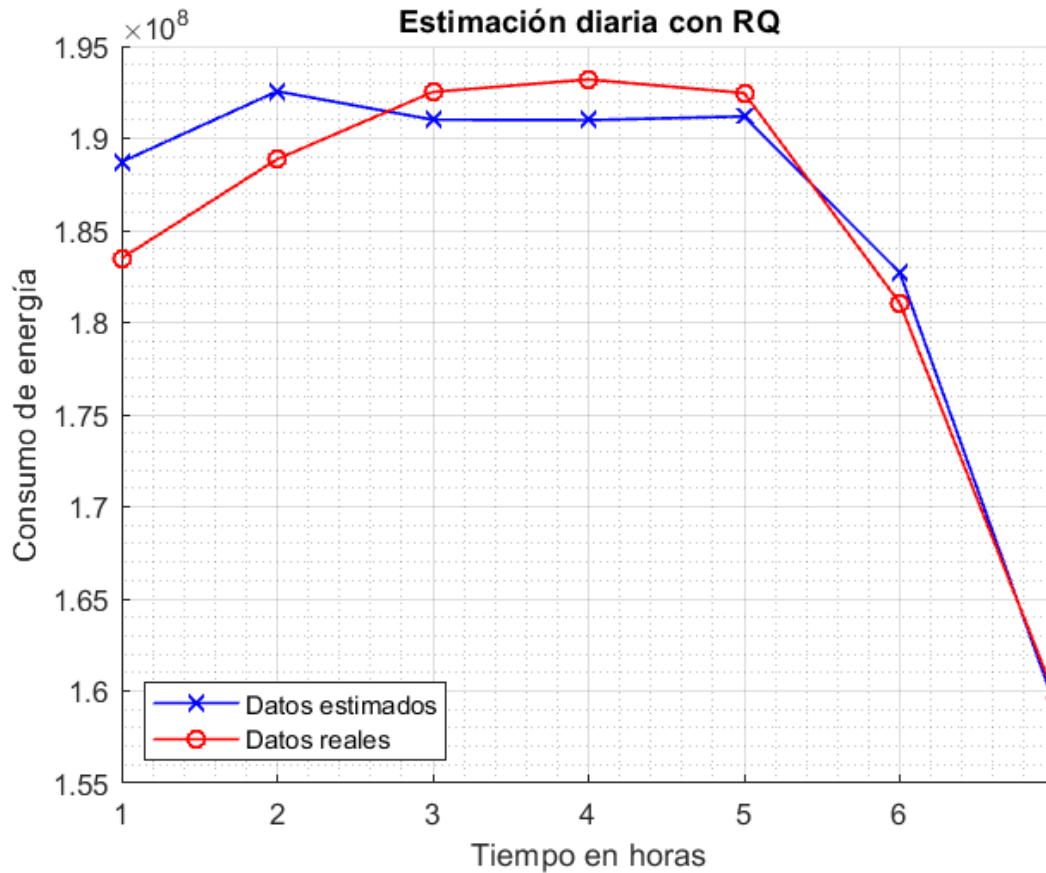


Figura 4.4: Estimación diaria con la función de covarianza base RQ. La línea roja representa el consumo real. La línea azul representa las estimaciones.

Fuente: Elaboración propia.

observamos que la función compuesta realiza un mejor trabajo que las básicas.

4.2. Resultados predicción horaria

A continuación se presentan los resultados de las estimaciones por horas para los lunes, martes, miércoles, jueves y viernes del año 2015 empleando el kernel SE+PER. Con el fin de resaltar los mejores desempeños del modelo, se decide tomar los cinco mejores resultados por día.

La tabla 4.3 informa que el día lunes es el que presenta menor rendimiento. Esto puede ser debido a la presencia de días festivos los cuales se

Tabla 4.3: *Resultados de la predicción horaria en el año 2015.*

Día	SMSE	MSLL
Lunes	0.0008 ± 0.00040	-3.2506 ± 0.1631
Martes	0.0003 ± 0.00011	-3.3259 ± 0.0928
Miércoles	0.0002 ± 0.00018	-3.3132 ± 0.0947
Jueves	0.0002 ± 0.00017	-3.3517 ± 0.0764
Viernes	0.0004 ± 0.00008	-3.3204 ± 0.0663

manifiestan más en los días en cuestión. En contraste, el día miércoles es el que presenta mejor rendimiento apesar de que tiene un resultado similar al día jueves, pues su desviación estándar es cero. Este buen comportamiento puede ser explicado por la nula presencia de días festivos en los datos de entrenamiento para este día.

El experimento se repite con algunas funciones de covarianza base para constatar el desempeño de la función compuesta. Dichas funciones base son: SE, PER y RQ. También se tomaron los mejores cinco resultados.

Como puede observarse en la tabla 4.4, los resultados con la función base SE son un poco mejores o iguales a los obtenidos con la función compuesta. Esto nos indica que la función PER en la composición no está realizando un trabajo significativo. Es posible, entonces, que el comportamiento periódico de los datos de entrenamiento no presente un patrón cíclico el cual la función PER pueda manejar, razón por la cual esta función arroja unas pobres estimaciones como se evidencia en la tabla en cuestión.

Tabla 4.4: *Resultados de las estimaciones con las funciones base.*

	SMSE		
	SE	PER	RQ
Lunes	0.00082 ± 0.00027	0.02218 ± 0.0014	0.0007 ± 0.00018
Martes	$0.00028 \pm 8.3666E-05$	0.02106 ± 0.0023	$0.00028 \pm 8.3666E-05$
Miércoles	0.0002 ± 0.00018	0.01648 ± 0.0016	$0.00022 \pm 4.47214E-05$
Jueves	0.00024 ± 0.00016	0.0174 ± 0.0010	0.00036 ± 0.00015
Viernes	0.00028 ± 0.00010	0.0193 ± 0.0023	$0.00042 \pm 8.3666E-05$

La tabla 4.5 permite apreciar la media y desviación estándar del error SMSE al repetir el experimento en los años 2013 y 2014 con la función

compuesta. De acuerdo a los resultados ofrecidos en esta tabla, podemos deducir que el modelo presenta un comportamiento similar y/o estable en los diferentes años de repetición del experimento.

Tabla 4.5: Media y desviación estándar

	SMSE
Lunes	$0,000333333 \pm 5,7735E-05$
Martes	$0,000233333 \pm 0,000152753$
Miércoles	$0,000266667 \pm 0,00011547$
Jueves	$0,000300000 \pm 0,000264575$
Viernes	$0,000433333 \pm 0,000152753$

Datos obtenidos entre los años 2013 y 2015

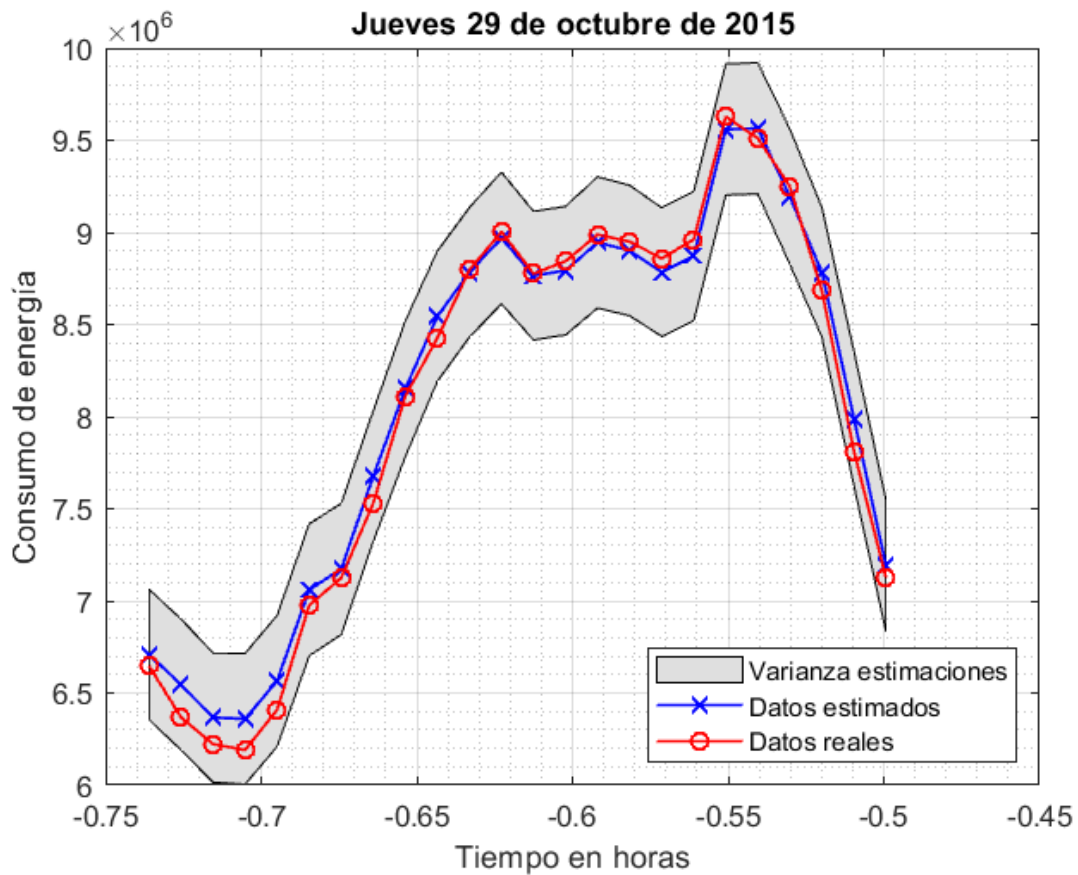


Figura 4.5: Predicción para el jueves 29 de octubre de 2015. Los datos en rojo indican el consumo real, los datos en azul indican las estimaciones.

Fuente: Elaboración propia.

El mejor ajuste se presenta para el día jueves 29 de octubre con un error SMSE de 0.0001. En los datos de entrenamiento se excluyó el día 1 de enero, pues este es un día que presenta demasiada regularidad en el consumo de energía por tratarse del primer día del año. La figura 4.2 nos ofrece una perspectiva gráfica de la predicción del jueves mencionado anteriormente con la variabilidad de las estimaciones representada en gris.

4.3. Métodos alternativos a los GP para predicción

En la sección 3.5, se mencionaron otros métodos de predicción, los cuales fueron comparados con los GP. A continuación, se relacionan sus parámetros y estructura empleados en la comparación.

Máquinas de vectores soporte

La cantidad de vectores de soporte en el entrenamiento es de 86 (Alpha) y el valor de ϵ que corresponde a la exactitud preestablecida de interpolación, es de $1,4635e + 05$.

Modelo ARIMA

Los parámetros p , d y q son números enteros no negativos que indican el orden de las distintas componentes del modelo, *autorregresiva*, *integrada* y de *media móvil*, respectivamente. Para este caso se empleó 26,1 y 1.

Artificial Neural Network (ANN)

En esta propuesta se empleó una NN con 48 datos de entrada, una capa oculta (hidden layer) de tamaño 10 y 24 capas de salida que representan las estimaciones.

Deep Neural Network (DNN)

Para la comparación horaria, se empleó una DNN [70] con cuatro capas y 48,20,50,1 neuronas por capa respectivamente.

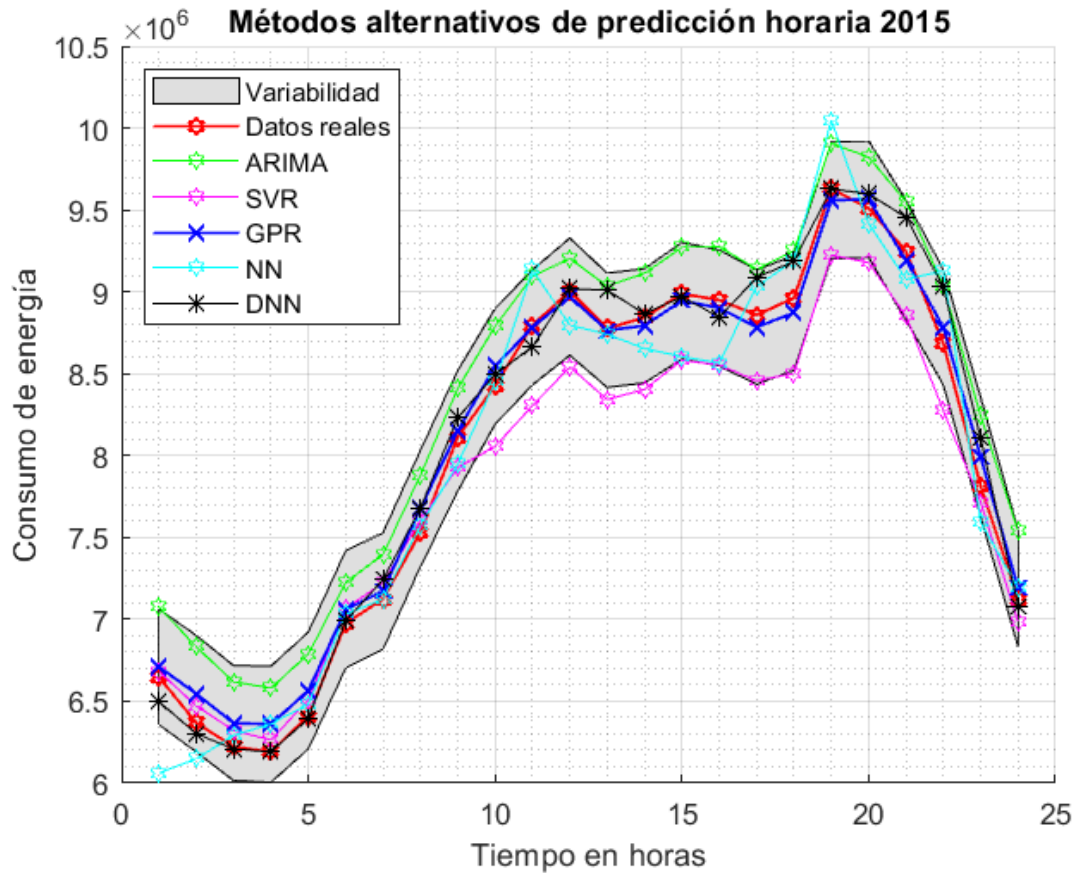


Figura 4.6: Predicción del jueves 29 de octubre 2015 con diferentes métodos de estimación. La línea roja representa el consumo real.

Fuente: Elaboración propia.

Tabla 4.6: Errores métodos alternativos horario entre los años 2013 y 2015

	SMSE	MSLL
ARIMA	0.0025 ± 0.0015	-2.9390 ± 0.3377
SVR	0.0012 ± 0.0003	-3.3257 ± 0.8850
GP	0.0003 ± 0.0002	-2.9908 ± 0.2730
NN	0.0042 ± 0.0027	-2.2816 ± 0.3526
DNN	0.0017 ± 0.0011	-3.2816 ± 0.4098

La figura 4.6 permite apreciar que el GP presenta el mejor rendimiento respecto a los otros métodos de predicción. En dicha figura podemos apreciar en gris la variabilidad o región de confianza que ofrecen los GP. Esta es una característica que permite tratar la incertidumbre inherente a

los datos por variables que no se pueden controlar o explicar. La tabla 4.6 corrobora la información suministrada por la mencionada figura.

4.4. Predicción de demanda usando diferentes conjuntos de datos

En esta sección se emplea un conjunto de datos diferentes para realizar predicciones usando los GP. Estos datos fueron empleados en un trabajo similar [71]. Los datos fueron aportados por Belgian Electricity Supplier Elia¹. Las predicciones se realizan con la función de covarianza compuesta SE+PER.

Estas predicciones pueden observarse en las figuras 4.4 y 4.5, donde se aprecian estimaciones horarias y semanales respectivamente.

¹<http://www.elia.be/>

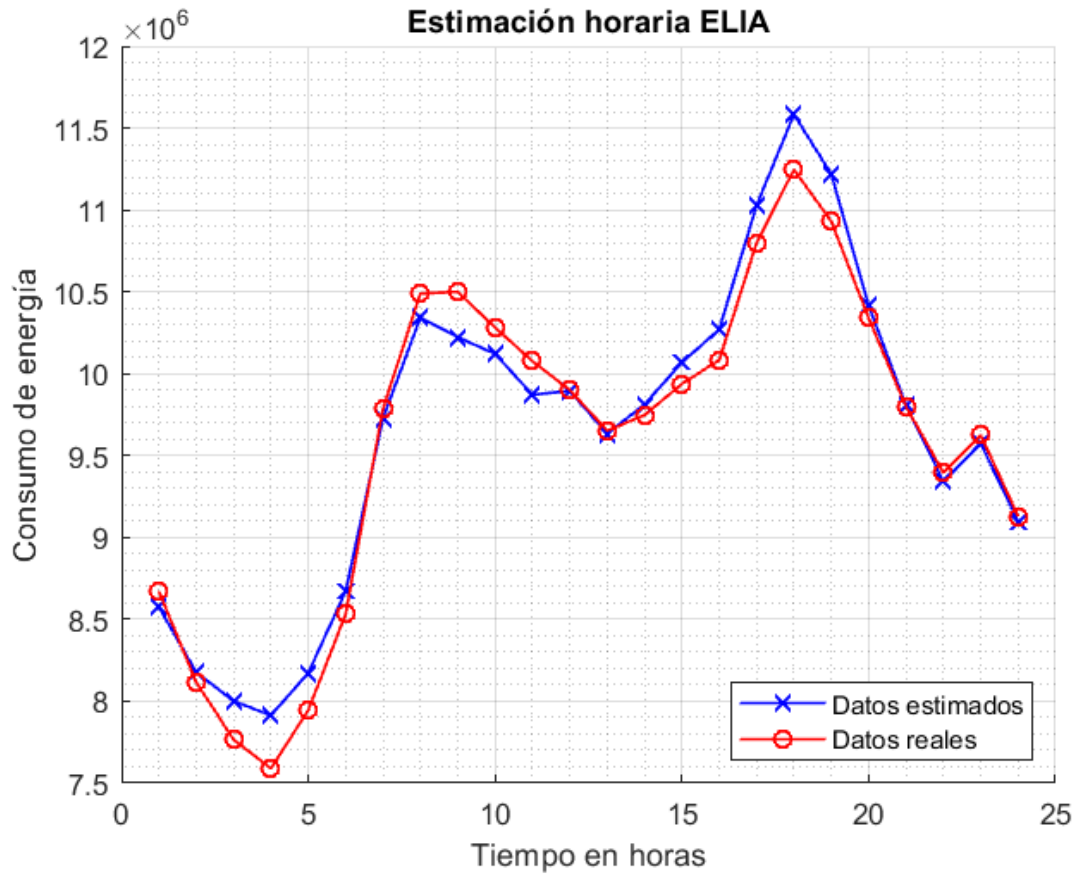


Figura 4.7: Estimaciones para el día jueves de 2015 con la función $SE+PER$. Los datos en rojo representan el consumo real, los azules las estimaciones

Fuente: Elaboración propia.

En la figura 4.7 se aprecia la mejor estimación horaria la cual corresponde al jueves número 12 de los datos de prueba y cuyo error SMSE es de 0.0003. De manera general los errores SMSE y MSL son respectivamente 0.008375 ± 0.01114 y -4.4304 ± 6.3446

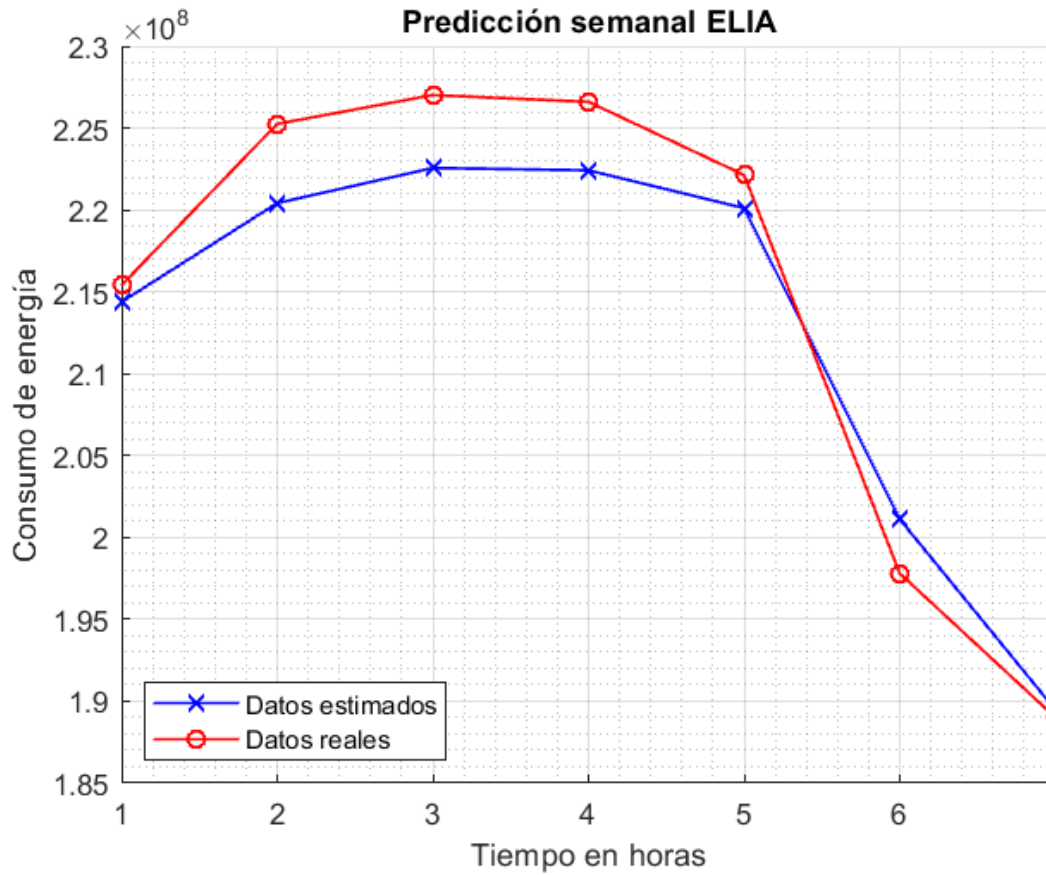


Figura 4.8: *Estimación semanal de los datos alternos. Los datos en rojo representan el consumo real, los azules las estimaciones.*

Fuente: Elaboración propia.

La fig. 4.8 permite apreciar el resultado de las estimaciones siguiendo la misma metodología para el horizonte semanal. La mejor estimación se presenta para la semana número 2 de los datos de prueba y presenta un error SMSE de 0.0002. Los resultados de manera general son 0.005496 ± 0.007369 para SMSE y -2.4186 ± 0.2611 para MSLL.

CONCLUSIONES

Los resultados obtenidos en esta propuesta de investigación permiten determinar que los GP, con la función de covarianza propuesta, son una buena alternativa para la predicción de series de tiempo. El buen desempeño se debe a que son flexibles, fáciles de implementar, poseen un enfoque probabilístico adecuado para un amplio rango de problemas en regresión y permiten estimar la incertidumbre de las predicciones. También al ser un modelo no-paramétrico, no está sujeto o restringido a ciertos valores en los parámetros de la función de covarianza (hyperparámetros), los cuales pueden ser optimizados por diferentes algoritmos.

Por otro lado, el éxito de un proceso gaussiano radica en la adecuada selección de la función de covarianza, pues ésta debe ir acorde al comportamiento que presenten los datos considerando las características propias de cada una de ellas. Es decir, el análisis previo de los datos permite observar la tendencia o comportamiento que estos presentan, dando así una idea del kernel que se puede emplear para su tratamiento. Sin embargo, en algunos casos los datos no presentan un comportamiento definido en su totalidad o solo parcialmente. Se hace necesario, entonces, formular o establecer nuevas funciones kernel que permitan procesar estos comportamientos irregulares.

Trabajos futuros

Como sugerencia para futuros trabajos, se recomienda implementar nuevas funciones de covarianza compuestas de acuerdo con el comportamiento

de los datos iniciales. Una buena opción podría ser una composición por multiplicación de SE y PER. También, puede realizarse una composición construyendo un árbol de búsqueda [51], sin dejar de lado la parte intuitiva del investigador pues por más herramientas diseñadas para ayudar en cualquier proceso, el análisis del autor jamás sobrar . En cuanto a los datos de entrenamiento, pueden emplearse otras heur sticas para tratar de mejorar las predicciones. Asi mismo, dado que los GP tienen un alto costo computacional, ser  pr ctico adaptar el modelo de esta propuesta a una mayor cantidad de datos de entrenamiento sin que ello signifique una p rdida de rendimiento o estabilidad en el proceso de estimaci n, es decir, ser  interesante escalar el modelo.

Por  ltimo, quedan otras t cnicas de estimaci n para comparar con los GPR tales como: modelos vectoriales autorregresivos (VAR), modelo de vectores de correcci n de error (VEC), algoritmos gen ticos [72], AdaBoost (ABR) [69], otros tipo de redes neuronales tales como Red de Hopfield, Red Competitiva, Redes Neuronales ART1, Red Neuronal ART2, Redes Neuronales Autorganizativas (Mapas de Kohonen), entre otras.

REFERENCIAS

- [1] DE LA POMPA PORRAS, Víctor. Procesos Gaussianos para problemas de regresión y estimación de la incertidumbre. UAM. Departamento de Ingeniería Informática (2018).
- [2] Adrián Ramírez del Río. (2015). Predicción de incertidumbres en demandas mediante Procesos Gaussianos (tesis de pregrado). Universidad Autónoma de Madrid, España.
- [3] Yang, Y., Li, S., Li, W., & Qu, M. (2018). Power load probability density forecasting using Gaussian process quantile regression. *Applied Energy*, 213, 499-509.
- [4] Bing Donga, Zhaoxuan Lia, S.M. Mahbobur Rahmana, Rolando Vega. A hybrid model approach for forecasting future residential electricity consumption. 17 June 2015. *Energy and Buildings* 117 (2016) 341–351.
- [5] Yulai Zhang, Guiming Luo, Fuan Pu. Power Load Forecasting based on Multi-task Gaussian Process. Funds NSFC61171121 and the Science Foundation of Chinese Ministry of Education - China Mobile 2012.
- [6] Lourenço J.M., Santos P.J. Short Term Load Forecasting Using Gaussian Process Models. June 8, 2010. Instituto de Engenharia de Sistemas e Computadores de Coimbra INESC – Coimbra.
- [7] Miltiadis Alamaniotis, Dimitrios Bargiotas, Lefteri H. Tsoukalas. Towards smart energy systems: application of kernel machine regression

- for medium term electricity load forecasting. *Alamaniotis et al. SpringerPlus* (2016) 5:58 DOI 10.1186/s40064-016-1665-z
- [8] Pramukpong Atsawathawichok, Prasit Teekaput, Tuchsana Ploysuwan. “Long term Peak Load Forecasting in Thailand using Multiple Kernel Gaussian Process”. 978-1-4799-2993-1/14/31.00 ©2014 IEEE.
 - [9] Fan G., Peng L., Hong W., (2018). Short term load forecasting based on phase space reconstruction algorithm and bi-square kernel regression model. *ScienceDirect*, 224, 13-33. <https://doi.org/10.1016/j.apenergy.2018.04.075>
 - [10] Van Der Meer D., Shepero M., Svensson A., Widén J., Munkhammar J. (2018). Probabilistic forecasting of electricity consumption, photovoltaic power generation and net demand of an individual building using Gaussian Process. *ScienceDirect*, 213, 195-207. <https://doi.org/10.1016/j.apenergy.2017.12.104>
 - [11] Zhang J., Wei Y., Li D., Tan Z., Zhou J. (2018). Short term electricity load forecasting using a hybrid model. *ScienceDirect*, 158, Pages 774-781. <https://doi.org/10.1016/j.energy.2018.06.012>
 - [12] Tarsitano A., Amerise I. (2017). Short term load forecasting using a two-stage sarimax model. *ScienceDirect*, 133, Pages 108-114. <https://doi.org/10.1016/j.energy.2017.05.126>
 - [13] Dudek G. (2016). Pattern based local linear regression models for short term load forecasting. *ScienceDirect*, 130, Pages 139-147. <https://doi.org/10.1016/j.epsr.2015.09.001>
 - [14] Zamo M. Mestre O., Arbogast P., Pannekouke O. (2014). A benchmark of statistical regression methods for short-term forecasting of photovoltaic electricity production. *ScienceDirect*, 105, Pages 792-803. <https://doi.org/10.1016/j.solener.2013.12.006>
 - [15] De Oliveira E., Oliveira F. (2018). Forecasting mid-long-term electricity energy consumption through bagging ARIMA and exponential smoothing methods. *ScienceDirect*, 144, Pages 776-788. <https://doi.org/10.1016/j.energy.2017.12.049>

- [16] Takeda H., Tamura Y., Sato S. (2016). Using the ensemble Kalman filter for electricity load forecasting and analysis. *ScienceDirect*, 104, Pages 184-198. <https://doi.org/10.1016/j.energy.2016.03.070>
- [17] Duan Q., Liu J., Zhao D. (2017) Shor-term electric load forecasting using an automated system of model choice. *ScienceDirect*, 91, Pages 92-100. <https://doi.org/10.1016/j.ijepes.2017.03.006>
- [18] Chen Y., Xu P., Chu Y., Li W., Ni L. (2017). Short-term electrical load forecasting using the support vector regression. *ScienceDirect*, 195, Pages 659-670. <https://doi.org/10.1016/j.apenergy.2017.03.034>
- [19] Fan G., Peng L., Hong W., Sun F. (2016). Electric load forecasting by the SVR model with differential empirical mode decomposition and auto regression. *ScienceDirect*, 173, Pages 958-970. <https://doi.org/10.1016/j.neucom.2015.08.051>
- [20] Geng J. Huang M., Li M., Hong W. (2015). Hybridization of seasonal chaotic cloud simulated annealing algorithm in a SVR-based load forecasting model. *ScienceDirect*, 151, Pages 1362-1373. <https://doi.org/10.1016/j.neucom.2014.10.055>
- [21] Sadaei H., Guimaraes F. da Silva C., Lee M. Eslami T. (2017). Short-term load forecasting method base in fuzzy time series, seasonality and long memory process. *ScienceDirect*, 83, Pages 196-217. <https://doi.org/10.1016/j.ijar.2017.01.006>
- [22] Singh P. Dwivedi P. (2018). Integration of new evolutionary approach with artificial neural network for solving short term load forecast problem. *ScienceDirect*, 217, Pages 537-549. <https://doi.org/10.1016/j.apenergy.2018.02.131>
- [23] Sheng H., Xiao J. Cheng Y., Ni Q., Wang S. (2017). Short-term solar power forecasting based on weighted Gaussian process regression. *IEEE Xplore*, 65, pages 300-308. DOI: 10.1109/TIE.2017.2714127
- [24] Lauret P., David M., Calogine D. (2012). Nonlinear models for short-time load forecasting. *ScienceDirect*, 14, Pages 1404-1409. <https://doi.org/10.1016/j.egypro.2011.12.1109>

- [25] Proyección de Demanda de Energía Eléctrica en Colombia. Revisión noviembre de 2014. UPME.
- [26] PROYECCIÓN DE LA DEMANDA DE ENERGÍA ELÉCTRICA Y POTENCIA MÁXIMA EN COLOMBIA. Revisión abril de 2018. UPME.
- [27] Stock, J.H. y M.W. Watson (2007), *Introduction to Econometrics*. 2da ed. Estados Unidos de América, Pearson, págs. 4-795.
- [28] María de la Paz Guzmán Plata, Pascual García Alba Iduñate. (2008). “EL MODELO VAR Y SUS PRINCIPALES PROBLEMAS”. *PANORAMA ECONÓMICO*, Volumen III, No. 6, enero-junio, 2008, pp. 95-117.
- [29] Guerrero, J. F. J., Fernández, R. S., Abad, J. C. G.: La capacidad predictiva en los métodos Box-Jenkins y Holt-Winters: una aplicación al sector turístico. *Revista Europea de Dirección y Economía de la Empresa*, 15(3), pp. 185–198 (2006).
- [30] James Robert Lloyd. (2013). “GEFCom2012 hierarchical load forecasting:” Gradient boosting machines and Gaussian processes”. *ELSEVIER, International Journal of Forecasting* 30 (2014) 369–374
- [31] JHOUBEN JANYK CUESTA RAMÍREZ. (2013). “ANÁLISIS DE EXPANSIÓN DE REDES DE TELEFONÍA MÓVIL BASÁNDOSE EN INDICADORES CLAVES DE DESEMPEÑO, UTILIZANDO PROCESOS GAUSSIANOS”. UTP, Pereira.
- [32] Sofiane Brahim-Belhouari, Amine Bermak. (2004). “Gaussian process for nonstationary time series prediction”. *ScienceDirect*, Volumen 47, Pages 705-712.
- [33] Ernesto Varela Arregocés y Edwin Campbells Sánchez. (2011). “Redes Neuronales Artificiales: Una Revisión del Estado del Arte, Aplicaciones Y Tendencias Futuras”. *Investigación y Desarrollo en TIC*, Vol 2 No 1 pp. 18-27.

- [34] Pai, P.; Hong, W. (2005). Forecasting regional electricity load based on recurrent support vector machines with genetic algorithms. *Electric Power Systems Research*, 74(3), 417–425.
- [35] Bunn, D.W.; Farmer, E. D. (1985). *Comparative models for electrical load forecasting*. New York: John Wiley.
- [36] Viviana María Rueda Mejía. (2011). “Predicción del Consumo de Energía en Colombia con Modelos no Lineales” (tesis de maestría). Universidad Nacional de Colombia, Medellín.
- [37] Velásquez, J.D.; Dyner, I.; Souza, R.C. (2007). ¿Por qué es tan difícil obtener buenos pronósticos de los precios de la electricidad en mercados competitivos? *Cuadernos de Administración*, 20, 34, 259 – 282.
- [38] Márquez Cebrián, M. D. "Modelo setar aplicado a la volatilidad de la rentabilidad de las acciones: algoritmos para su identificación". Tesis doctoral, UPC, Departament d'Estadística i Investigació Operativa, 2002.
- [39] Rueda, V. M., HENAO, J. D. V., & CARDONA, C. J. F. (2011). Avances recientes en la predicción de la demanda de electricidad usando modelos no lineales. *Dyna*, 78(167), 36-43.
- [40] Gutiérrez, E. (2003). La demanda residencial de energía eléctrica en la Comunidad Autónoma de Andalucía: un análisis cuantitativo. Tesis de Doctorado. Facultad de Ciencias Económicas y empresariales, Universidad de Sevilla.
- [41] Orellana Romero, J. L. (2012). Modelación y pronóstico de la demanda de energía eléctrica de mediano plazo de El Salvador (Doctoral dissertation, Universidad de El Salvador).
- [42] S. M. Al-Alawi and S. M. Islam, "Principles of electricity demand forecasting. I. Methodologies," in *Power Engineering Journal*, vol. 10, no. 3, pp. 139-143, June 1996. doi: 10.1049/pe:19960306

- [43] Medina, S.; García, J. (2005). Predicción de demanda de energía en Colombia mediante un sistema de inferencia difuso neuronal. *Revista Energética*, 33, 15–24.
- [44] Valencia, A.L.; Lozano, C.A.; Moreno, C.A. (2007) Modelo de promedios móviles para el pronóstico horario de potencia y energía eléctrica. *El Hombre y la Máquina*, 29, 96 - 105
- [45] Murillo, J.; Tejos, A.; Carvajal, P. (2003). Estudio del pronóstico de la demanda de energía eléctrica utilizando modelos de series de tiempo. *Scientia et Technica*, 23, 37–42.
- [46] C. E. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning*. The MIT Press, 2006.
- [47] K. Chalupka, “Empirical evaluation of gaussian process approximation algorithms,” Ph.D. dissertation, School of Informatics, University of Edinburgh, 2011.
- [48] W. Al-Hawani, M. Samarasinghe,” Short-term Forecasting of Electricity Consumption using Gaussian Processes,” master’s thesis, University of Agder, 2012.
- [49] Kevin P. Murphy. *Machine Learning: A Probabilistic Perspective*. Adaptive computation and machine learning series. MIT Press, Cambridge, MA, 2012. ISBN 978-0-262-01802-9.
- [50] Wilson, A., & Adams, R. (2013, February). Gaussian process kernels for pattern discovery and extrapolation. In *International Conference on Machine Learning* (pp. 1067-1075).
- [51] Bagas Abisena, Swastanto. (2016). “Gaussian Process Regression for Long-Term Time Series Forecasting”. Delft University of Technology.
- [52] Jhon Cuellar, "Pronóstico de la Demanda de Energía Eléctrica Empleando Procesos Gaussianos con el Kernel Mezcla Espectral", Ing. tesis, Dept Elect., UTP. 2016.
- [53] LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4), 541-551.

- [54] Poon, H., & Domingos, P. (2011, November). Sum-product networks: A new deep architecture. In 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops) (pp. 689-690). IEEE.
- [55] David K. Duvenaud, James Robert Lloyd, Roger B Grosse, Joshua B Tenenbaum, and Zoubin Ghahramani. Structure Discovery in Nonparametric Regression through Compositional Kernel Search. In ICML (3), pages 1166–1174, 2013.
- [56] Pasan Karunaratne, Masud Moshtaghi, Shanika Karunasekera, Aaron Harwood, Trevor Cohn. “Modelling the Working Week for Multi-Step Forecasting using Gaussian Process Regression”. Department of Computing and Information Systems, The University of Melbourne, Australia.
- [57] Carl Edward Rasmussen y Hannes Nickisch. (2018). “The GPML Toolbox version 4.2”.
- [58] David Duvenaud, James Robert Lloyd, Roger Grosse, Joshua B. Tenenbaum, Zoubin Ghahramani. (2013).” Structure Discovery in Nonparametric Regression through Compositional Kernel Search”. arXiv:1302.4922v4 [stat.ML].
- [59] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3), 1995.
- [60] D. Li, W. Xu, H. Zhao and Rong-Qiu Chen, .^A SVR based forecasting approach for real estate price prediction", 2009 International Conference on Machine Learning and Cybernetics, Hebei, 2009, pp. 970-974.doi: 10.1109/ICMLC.2009.5212389
- [61] Enrique J. Carmona Suárez. “Tutorial sobre Máquinas de Vectores Soporte (SVM)”, Dpto. de Inteligencia Artificial, ETS de Ingeniería Informática, Universidad Nacional de Educación a Distancia (UNED), C/Juan del Rosal, 16, 28040-Madrid (Spain).
- [62] Ojemakinde, Bukola Titilayo, "Support Vector Regression for Non-Stationary Time Series. "Master's Thesis, University of Tennessee, 2006. <https://trace.tennessee.edu/utkgradthes/1756>

- [63] D. Wu, "Time Series Prediction for Machining Errors Using Support Vector Regression," 2008 First International Conference on Intelligent Networks and Intelligent Systems, Wuhan, 2008, pp. 27-30. doi: 10.1109/ICINIS.2008.31
- [64] S. A. Dyer and Xin He, "Least-squares fitting of data by polynomials," in IEEE Instrumentation & Measurement Magazine, vol. 4, no. 4, pp. 46-51, Dec. 2001. doi: 10.1109/5289.975465
- [65] António Araújo, "Polynomial regression with reduced overfitting—The PALS technique", Volume 124, August 2018, Pages 515-521. In Science Direct.
- [66] G. M. U. Din and A. K. Marnerides, "Short term power load forecasting using Deep Neural Networks," 2017 International Conference on Computing, Networking and Communications (ICNC), Santa Clara, CA, 2017, pp. 594-598.
- [67] X. Hui, W. Qun, L. Yao, Z. Yingbin, S. Lei and Z. Zhisheng, "Short-term load forecasting model based on deep neural network," 2017 2nd International Conference on Power and Renewable Energy (ICPRE), Chengdu, 2017, pp. 589-591.
- [68] Carl Edward Rasmussen. (1997). "Evaluation of Gaussian Processes and Other Methods for Non-Linear Regression". Department of Computer Science in the University of Toronto Convocation of March 1997.
- [69] Adrián Ramírez del Río. (2015). Predicción de incertidumbres en demandas mediante Procesos Gaussianos (tesis de pregrado). Universidad Autónoma de Madrid, España.
- [70] Masayuki Tanaka and Masatoshi Okutomi, A Novel Inference of a Restricted Boltzmann Machine, International Conference on Pattern Recognition (ICPR2014), August 2014.
- [71] Jaehwa Lee, Pengfei Zhang, Leong Kit Gan, David A. Howey, Michael A. Osborne, Alessandra Tosi y Stephen Duncan. (2018). Optimal operation of an energy management system using model predictive control

- and Gaussian process time-series modelling. *IEEE Journal of Emerging and Selected Topics in Power Electronics*, 6(4), 1783 – 1795.
- [72] Niu Dongxiao, Li Jinchao, Li Jinying y Liu Da. (2007). Middle-long Power Load Forecasting Based on Genetic Algorithm. *IEEE*, 26-31 July 2007.